ESAIM: M2AN 54 (2020) 1309–1337
https://doi.org/10.1051/m2an/2019047

ESAIM: Mathematical Modelling and Numerical Analysis
www.esaim-m2an.org

RECOVERED FINITE ELEMENT METHODS ON POLYGONAL AND POLYHEDRAL MESHES

Zhaonan Dong[1], Emmanuil H. Georgoulis[2,3] and Tristan Pryer[4,*]

Abstract. *Recovered Finite Element Methods (R-FEM)* have been recently introduced in Georgoulis and Pryer [*Comput. Methods Appl. Mech. Eng.* **332** (2018) 303–324]. for meshes consisting of simplicial and/or box-type elements. Here, utilising the flexibility of the R-FEM framework, we extend their definition to polygonal and polyhedral meshes in two and three spatial dimensions, respectively. An attractive feature of this framework is its ability to produce arbitrary order polynomial *conforming* discretizations, yet involving *only* as many degrees of freedom as discontinuous Galerkin methods over general polygonal/polyhedral meshes with potentially many faces per element. *A priori* error bounds are shown for general linear, possibly degenerate, second order advection-diffusion-reaction boundary value problems. A series of numerical experiments highlight the good practical performance of the proposed numerical framework.

Mathematics Subject Classification. 65N30, 65N50, 65N55.

Received July 12, 2018. Accepted June 27, 2019.

1. Introduction

In recent years, we have witnessed considerable interest in the construction of Galerkin-type numerical methods over meshes consisting of general polygons in two dimensions or general polyhedra in three dimensions, henceforth termed collectively as *polytopic*, as opposed to the classical Galerkin methods employing simplicial and/or box-type meshes. This interest is motivated by the expectation that such generality of meshes can yield minimally invasive numerical coarse-graining, by representing exactly complicated geometries with fewer numerical degrees of freedom, compared to standard methods, and also for their potential use within adaptive algorithms involving both coarsening and refinement. The latter is particularly pertinent in the context of adaptive computations for evolution PDE problems, where dynamic mesh modification is widely accepted as a promising tool for the reduction of computational complexity in both Eulerian and Lagrangian contexts. Galerkin procedures over polytopic meshes have been also proposed in the context of interface problems

Keywords and phrases. Recovered finite element method, polygonal and polyhedral meshes, *a priori* analysis, PDEs with non-negative characteristic form.

[1] School of Mathematics, Cardiff University, Cardiff CF24 4AG, UK.

[2] Department of Mathematics, University of Leicester, University Road, Leicester LE1 7RH, UK.

[3] School of Applied Mathematical and Physical Sciences, National Technical University of Athens, Zografou 15780, Greece.

[4] Department of Mathematical Sciences, University of Bath, Claverton Down, Bath BA2 7AY, UK.

*Corresponding author: tmp38@bath.ac.uk, t.pryer@reading.ac.uk

(porosity profiles, interfaces, etc.), as well as a flexible method for coarse-scale correction computations in multilevel solvers for elliptic boundary-value problems.

Popular polytopic methods include the virtual element method [3, 14], which is itself an evolution of the so-called mimetic finite difference methods [4], polygonal finite element methods [41], BEM-based FEMs [39], composite finite element methods [28, 37] and various discontinuous Galerkin (dG) approaches, ranging from one-field interior penalty dG methods [10–12, 20], to hybridized formulations [18, 19, 21, 33] and discontinuous Petrov–Galerkin schemes [1]. Interior penalty DG methods (and related one-field formulations) allow for the control of the number the global numerical degrees of freedom independently of the mesh topology (*i.e.*, the connectivity of the nodes/faces/elements), whereas polygonal finite elements, virtual element methods and hybridized formulations are defined on local approximation spaces whose dimension depends on the number of faces/vertices for each polytopic element.

In this work, motivated by the recent recovered finite element framework presented in [24], we construct *conforming* discretizations over polytopic meshes whose set of degrees of freedom is *independent* of the number of vertices/edges/faces of each element. The proposed family of methods, termed collectively as recovered finite element methods (R-FEMs) depends, instead, on the choice of a sub-triangulation of the polytopic meshes. Crucially, however, the computational complexity of the method is *independent* of the cardinality of the simplices in the sub-triangulation: R-FEM on polytopic meshes combines completely discontinuous local *polynomial* spaces, resulting, nonetheless, to conforming approximations.

To fix ideas, let us consider an elliptic boundary value problem with homogeneous Dirichlet boundary conditions. Let $\mathcal{E} : V_h \to \widetilde{V}_h \cap H_0^1(\Omega)$ an operator mapping a *discontinuous* element-wise polynomial space V_h over a polytopic mesh onto a space of *continuous* piecewise polynomial space $\widetilde{V}_h \cap H_0^1(\Omega)$ over a, generally speaking, finer simplicial mesh arising from a sub-triangulation of the polytopic mesh; such *recovery* operators $\mathcal{E}$ can be constructed locally, *e.g.*, by (weighted) averaging of the nodal degrees of freedom [16, 32, 43]. We can now consider the method: find $u_h \in V_h$, such that

$$\int_\Omega \nabla\mathcal{E}\left(u_h\right) \cdot \nabla\mathcal{E}\left(v_h\right) \mathrm{d}x + s\left(u_h, v_h\right) = \int_\Omega f\mathcal{E}\left(v_h\right) \mathrm{d}x, \quad \text{for all} \quad v_h \in V_h,$$

for $f \in H^{-1}(\Omega)$ and a suitable *stabilization* $s(\cdot,\cdot) : V_h \times V_h \to \mathbb{R}$, whose functionality is the treatment of the kernel $\{0 \neq v_h \in V_h : \mathcal{E}\left(v_h\right) = 0\}$ to achieve unisolvence. Notice that the method produces both a conforming approximation $\mathcal{E}\left(u_h\right)$ along with the non-conforming u_h, in spite of using element-wise discontinuous polynomial trial and test space V_h. In the limit case of the above R-FEM posed on a simplicial mesh (rather than a general polytopic one), $\mathcal{E}\left(u_h\right)$ corresponds to the classical conforming FEM approximations for certain choices of $\mathcal{E}$; we refer to [24] for details. Therefore, in this sense, R-FEM is an extension of classical finite element methods to polytopic meshes. An interesting property of the proposed method is that the user has access to the computed approximate solution at every point in the computational domain. This may be of practical interest both in the context of further post-processing and in the visualisation of the computation on standard widely available software.

As we shall see below, although sharing the same nominal complexity, R-FEMs introduced in this work will require more restrictive assumptions on the polytopic meshes for the respective error analysis to hold, compared to the current level of development of interior penalty DG methods on polytopic meshes [10–12]. Nonetheless, the R-FEM framework is envisaged to allow for a number of potentially attractive attributes compared to DG methods in designing compatible discretizations, preserving certain properties of the respective exact solutions at the discrete level. This is, precisely, due to the fact that, in R-FEMs, the conforming discrete solution $\mathcal{E}\left(V_h\right)$ is subordinate to the variational structure of the original PDE and, therefore, imitate (or even retain) various properties of the exact problem. Such compatibility considerations are particularly pertinent in nonlinear PDE problems or systems, requiring, *e.g.*, monotonicity, positivity preservation, etc. At the same time, since $\{0 \neq v_h \in V_h : \mathcal{E}\left(v_h\right) = 0\}$ is allowed to be *non-trivial* by construction, R-FEM offers significant flexibility by naturally separating the conformity-compatibility requirements of each problem, embedded into the choice of recovery operators, from the approximation properties of the local finite element spaces, thus

allowing for optimal approximation in very general numerical degrees of freedom scenarios, *e.g.*, various types of numerical degrees of freedom (nodal, modal, moments, etc.) for different elements in the same mesh. This can be, for instance, of interest in the consistent treatment of bulk/interface variational problems. Moreover, the uncoupling of conformity/compatibility from the local approximation space dimensions, can allow for *new* stable pairs for mixed R-FEM formulations; this will be discussed in detail in [2].

As a first step towards this programme, in the present work we restrict the discussion to linear PDEs with non-negative characteristic form, aiming to study the basic principles of construction of R-FEMs for polytopic meshes, firstly in an abstract coercive setting and secondly in for an example of an R-FEM requiring the proof of a discrete inf-sup condition for stability. As such, we leave the design of property/compatibility-preserving R-FEMs for particular nonlinear problems as a future challenge.

The remainder of this work is structured as follows. In Section 2, we introduce the problem and define a set of polytopic meshes. In Section 3, we introduce the FEM spaces and the recovery operators. Section 4 presents the concepts and ideas for designing R-FEM. In Section 5, we define the R-FEM for the model problem. The *a priori* error analysis for R-FEM is presented in Section 6. Finally, the practical performance of the proposed R-FEM is tested through a series of numerical examples in Section 7.

2. Model problem

Throughout this work we denote the standard Lebesgue spaces by $L^p(\omega)$, $1 \leq p \leq \infty$, $\omega \subset \mathbb{R}^d$, $d = 2, 3$, with corresponding norms $\|\cdot\|_{L^p(\omega)}$; the norm of $L^2(\omega)$ will be denoted by $\|\cdot\|_\omega$ for brevity. Let also $W^{s,p}(\omega)$ and $H^s(\omega) := W^{s,2}(\Omega)$, be the Banach and Hilbertian Sobolev space of index $s \in \mathbb{R}$ of real-valued functions defined on $\omega \subset \mathbb{R}^d$, respectively, constructed *via* standard interpolation and/or duality procedures for $s \notin \mathbb{N}_0$. For $H^s(\omega)$, we denote the corresponding norm and seminorm by $\|\cdot\|_{s,\omega}$ and $|\cdot|_{s,\omega}$, respectively. We also denote by $H^1_0(\omega)$ the space of functions in $H^1(\omega)$ with vanishing trace on $\partial\omega$.

Let Ω be a bounded open polygonal domain in $\mathbb{R}^d$, $d \in \mathbb{N}$, with $\partial\Omega$ denoting its boundary. We consider advection-diffusion-reaction problem

$$\mathcal{L}u \equiv -\nabla \cdot (a\nabla u) + \mathbf{b} \cdot \nabla u + cu = f \quad \text{in} \quad \Omega, \tag{2.1}$$

where $f \in L^2(\Omega)$, $\mathbf{b} \in \left[W^{1,\infty}(\Omega)\right]^d$, $c \in L^\infty(\Omega)$, for some definite diffusion tensor $a \in [L^\infty(\Omega)]^{d\times d}$ satisfying

$$\zeta^T a(x)\zeta \geq 0, \quad \text{for all} \quad \zeta \in \mathbb{R}^d, \tag{2.2}$$

for almost every $x \in \bar{\Omega}$. This class of problem is often termed *PDEs with non-negative characteristic form* [35] and includes elliptic, parabolic, first order hyperbolic as well as other non-standard types of PDEs, such as ultra-parabolic and various classes of linear degenerate equations. In particular, the important family of linear Kolmogorov–Fokker–Planck equations are of the form (2.1).

To prescribe suitable boundary conditions, we begin by splitting $\partial\Omega$ into

$$\Gamma_0 := \left\{x \in \partial\Omega : \mathbf{n}^T(x)a(x)\mathbf{n}(x) > 0\right\},$$

and

$$\Gamma_1 := \left\{x \in \partial\Omega : \mathbf{n}^T(x)a(x)\mathbf{n}(x) = 0\right\},$$

with $\mathbf{n}(x)$ denoting the unit outward normal vector to Ω at $x \in \partial\Omega$; the latter is further subdivided into inflow

$$\Gamma_- := \{x \in \Gamma_1 : \mathbf{b}(x) \cdot \mathbf{n}(x) < 0\},$$

and outflow $\Gamma_+ := \Gamma_1 \backslash \Gamma_-$ parts of the boundary. The "elliptic" part of the boundary Γ_0, is subdivided into Γ_D and Γ_N, on which we can prescribe Dirichlet and Neumann boundary conditions, respectively. For simplicity,

we assume that $|\Gamma_{\mathrm{D}}| > 0$, with $|\cdot|$ denoting the Hausdorff measure with the dimension of its argument, and that $\mathbf{b}(x)\cdot\mathbf{n}(x) \geq 0$ for almost all $x \in \Gamma_{\mathrm{N}}$. To complete the problem, we impose the boundary conditions:

$$\begin{aligned} u &= g_{\mathrm{D}}, \quad \text{on} \quad \Gamma_{\mathrm{D}} \cup \Gamma_{-}, \\ (a\nabla u)\cdot\mathbf{n} &= g_{\mathrm{N}}, \quad \text{on} \quad \Gamma_{\mathrm{N}}, \end{aligned} \tag{2.3}$$

for some known $g_{\mathrm{D}} \in L^2(\partial\Omega)$ and $g_{\mathrm{N}} \in H^{1/2}(\partial\Omega)$. For convenience, we also define the set

$$\Gamma_{\mathrm{D}}^{-} := \{x \in \partial\Omega : \mathbf{b}(x)\cdot\mathbf{n}(x) < 0\},$$

i.e., the inflow part of the boundary including also, possibly, points of Γ_{D}. Similarly, we define $\Gamma_{\mathrm{N}}^{+} = \partial\Omega\backslash\Gamma_{\mathrm{D}}^{-}$. Additionally, we assume that the following positivity hypothesis holds: there exists a positive constant γ_0 such that

$$c_0(x) := c(x) - \frac{1}{2}\nabla\cdot\mathbf{b}(x) \geq \gamma_0 \quad \text{a.e.} \quad x \in \Omega. \tag{2.4}$$

The well-posedness of the boundary value problem (2.1), (2.3) has been studied in [29].

3. Finite element spaces and recovery operators

Let $\mathcal{T}$ be a subdivision of Ω into disjoint polygonal elements for $d = 2$ or to disjoint polyhedral elements for $d = 3$; henceforth, these will be collectively referred to as *polytopic elements*. For simplicity, we assume that the subdivision $\mathcal{T}$ can be further subdivided into a conforming (*i.e.*, no hanging nodes) and shape-regular simplicial triangulation $\widetilde{\mathcal{T}}$ (see *e.g.*, [15], p. 124), that $\bar{\Omega} = \cup_{T\in\mathcal{T}}\bar{T}$. Such a setting can be constructed, *e.g.*, by agglomerating simplicial elements into polytopic ones.

By Γ we shall denote the union of all $(d-1)$-dimensional faces associated with the subdivision $\mathcal{T}$ including the boundary. Further, we set $\Gamma_{\mathrm{int}} := \Gamma\backslash\partial\Omega$. Correspondingly, we define $\widetilde{\Gamma}$ and $\widetilde{\Gamma}_{\mathrm{int}}$ for $\widetilde{\mathcal{T}}$. Note that, by construction, $\Gamma \subset \widetilde{\Gamma}$ and $\Gamma_{\mathrm{int}} \subset \widetilde{\Gamma}_{\mathrm{int}}$.

For a nonnegative integer r, we denote the set of all polynomials of total degree at most r by $\mathcal{P}_r(T)$. For $r \geq 1$, we consider the finite element space

$$V_h^r := \left\{v \in L^2(\Omega) : v|_T \in \mathcal{P}_r(T),\, T \in \mathcal{T}\right\}. \tag{3.1}$$

We stress that V_h^r is element-wise discontinuous polynomial with respect to the *polytopic* mesh $\mathcal{T}$; in this context, the dimension of V_h^r coincides with the dimension of discontinuous Galerkin finite element spaces on polytopic meshes, *cf.*, [10, 11, 13]. In particular, the dimension of V_h^r is *not* dependent on the number of vertices of the mesh $\mathcal{T}$. Correspondingly, we define

$$\widetilde{V}_h^r := \left\{v \in L^2(\Omega) : v|_T \in \mathcal{P}_r(T),\, T \in \widetilde{\mathcal{T}}\right\},$$

the respective discontinuous polynomial space on the sub-triangulation $\widetilde{\mathcal{T}}$. Note that $V_h^r \subset \widetilde{V}_h^r$. Moreover, we define the broken Sobolev space $H^1(\Omega,\mathcal{T})$ with respect to the subdivision $\mathcal{T}$ as follows:

$$H^1(\Omega,\mathcal{T}) := \left\{u \in L^2(\Omega) : u|_T \in H^1(T),\, T \in \mathcal{T}\right\}.$$

Further, let $T^+, T^- \in \mathcal{T}$ be two (generic) elements sharing a facet $e := \partial T^+ \cap \partial T^- \subset \Gamma_{\mathrm{int}}$ with respective outward normal unit vectors $\mathbf{n}^+$ and $\mathbf{n}^-$ on e. For a function $v : \Omega \to \mathbb{R}$ that may be discontinuous across Γ_{int}, we set $v^+ := v|_{e\subset\partial T^+}$, $v^- := v|_{e\subset\partial T^-}$, and we define the jump and average by

$$[v] := v^+\mathbf{n}^+ + v^-\mathbf{n}^- \quad \text{and} \quad \{v\} := \frac{1}{2}\left(v^+ + v^-\right);$$

if $e \in \partial T \cap \partial\Omega$, we set $[v] := v^+\mathbf{n}$. Also, we define $h_T := \mathrm{diam}(T)$ and we set $\mathbf{h} : \Omega\backslash\Gamma \to \mathbb{R}$, with $\mathbf{h}|_T = h_T$, $T \in \mathcal{T}$, $e \subset \Gamma_{\mathrm{int}}$ and $\mathbf{h}|_e = h_T$ for $e \subset \partial T \cap \partial\Omega$. Similarly, we set $\widetilde{\mathbf{h}}$ for the meshsize function of $\widetilde{\mathcal{T}}$. Throughout this work, we assume that the families of meshes are locally quasi-uniform and that there exists constant $c_\Delta > 1$, independent of the meshsizes such that

$$c_\Delta^{-1}\mathbf{h} \le \widetilde{\mathbf{h}} \le c_\Delta \mathbf{h},$$

uniformly as $\mathbf{h} \to 0$. Moreover, for the restriction of a function v on an element $T \in \mathcal{T}$, $v|_T : T \to \mathbb{R}$, which may be discontinuous across ∂T, we shall use the notational convention that $v^+|_{\partial T}$ signifies the trace from within T while $v^-|_{\partial T}$ signifies the trace from within $\Omega\backslash T$. Using this convention we also define the *signed jump* (also known as *upwind* jump in the discontinuous Galerkin literature) on each face e by

$$\lfloor v \rfloor|_e := v^+|_e - v^-|_e;$$

note that $|\,[v]\,| = |\lfloor v \rfloor|$.

Also, we shall denote by $\partial_- T$ and by $\partial_+ T$ the *inflow* and *outflow* parts of the boundary of an element T, defined as

$$\partial_- T := \{x \in \partial T : \mathbf{b}(x)\cdot\mathbf{n}(x) < 0\} \quad \text{and} \quad \partial_+ T := \{x \in \partial T : \mathbf{b}(x)\cdot\mathbf{n}(x) > 0\},$$

respectively.

For the definition of the proposed method, we shall require a *recovery operator* of the form

$$\mathcal{E} : V_h^r \to V \cap \widetilde{V}_h^r, \tag{3.2}$$

for some non-negative integer r, mapping element-wise discontinuous functions into functions in the solution space for the boundary value problem V, for some $r \in \mathbb{N}_0$. When the diffusion tensor a is strictly positive definite (*i.e.*, when (2.2) holds with strict inequality) we may take $V = H^1(\Omega)$.

Recovery operators of the form (3.2) have appeared in various settings in the theory of finite element methods, *e.g.*, [6, 16, 25, 32, 36, 40]. They are typically used to recover a "conforming" function from a "non-conforming" one, often under minimal regularity requirements.

A popular and very practical example for $\mathcal{E}$ is the nodal *averaging operator* for which the following celebrated stability result was proven by Karakashian and Pascal in [32].

Lemma 3.1. *Let $\mathcal{T}$ a polytopic mesh and $\widetilde{\mathcal{T}}$ its related sub-triangulation satisfying the above assumptions. Denoting by $\mathcal{N}$ the set of all Lagrange nodes of $\widetilde{V}_h^r \cap H^1(\Omega)$, the operator $\mathcal{E}_r : V_h^r \to \widetilde{V}_h^r \cap H^1(\Omega)$ is defined by:*

$$\mathcal{E}_r(v)(\nu) := \frac{1}{|\omega_\nu|}\sum_{T\in\omega_\nu} v|_T(\nu),$$

with $\omega_\nu := \bigcup_{T\in\mathcal{T}:\nu\in\bar{T}} T$, the set of elements sharing the node $\nu \in \mathcal{N}$ and $|\omega_\nu|$ their cardinality. Then, the following bound holds

$$\sum_{T\in\mathcal{T}} |v - \mathcal{E}_r(v)|^2_{\alpha,T} \le C_\alpha \left\|\mathbf{h}^{1/2-\alpha}\,[v]\right\|^2_{\Gamma_{\mathrm{int}}}, \tag{3.3}$$

with $\alpha \in \mathbb{N}_0$, $C_{|\alpha|} \equiv C_{|\alpha|}(r) > 0$ a constant independent of $\mathbf{h}$, v and $\widetilde{\mathcal{T}}$, but depending on the shape-regularity of $\widetilde{\mathcal{T}}$, on c_Δ, and on the polynomial degree r.

Proof. See Karakashian and Pascal [32]. □

The bound (3.3) shows, in particular, that $\left\|\mathbf{h}^{-1/2}\,[v]\right\|^2_{\Gamma_{\mathrm{int}}}$ is a norm on the orthogonal complement W_h^r of $\widetilde{V}_h^r \cap H^1(\Omega)$ in V_h^r with respect to the standard H^1-inner product.

It is possible to make a number of different choices

$$\mathcal{E} : V_h^r \to V \cap \widehat{V}_h^s, \tag{3.4}$$

for instance, for $\widehat{V}_h^s$ say a non-conforming finite element space, *e.g.*, Crouzeix–Raviart elements and s may be in general different to r. Indeed, various choices of $\mathcal{E}$ may give rise to different methods. As the present work focuses on the development of conforming methods on polytopic meshes, we prefer to keep the presentation simple and consider recovery operators into conforming element-wise polynomials of the same order ($r = s$) but, crucially, posed on different meshes. For an investigation of the case $r \neq s$ on standard element shapes, we refer to [24].

Remark 3.2. The proposed R-FEM scheme depends on the recovery operator which maps piecewise discontinuous polynomials defined on the polytopic elements to the conforming piecewise polynomials defined on the simplicial sub-meshes. The current mesh assumptions on the existence of a shape-regular sub-triangulation are necessitated to assert the validity of the recovery stability result Lemma 3.1. It would be an important direction of future research to investigate different stabilization terms, which give rise to norms in the recovery kernel (other than the standard jump penalty) that, in turn, will allow for small faces to be incorporated in the theory, to allow for mesh generality in the lines of other polytopic Galerkin methods, *e.g.*, [5, 7, 10–12]. Nevertheless, R-FEM works for very general polytopic meshes: in Section 7, we test the convergence of R-FEM for a numerical example (Example 5) on meshes designed by agglomerating much finer triangular meshes, resulting to many small faces per element.

For the error analysis below, we require an extension of the domain of definition of the recovery operator to

$$\mathcal{E} : \widetilde{V}_h^r \to V \cap \widehat{V}_h^s;$$

for instance, in the case of the averaging operator of Lemma 3.1 the extension is trivial and is given by the same formula.

4. Design concepts for recovered finite element methods

Equipped with a finite element space framework and the concept of recovery operators, we can now describe some general principles in the design of recovered finite element methods on polytopic meshes.

To this end, we consider a generic conforming Galerkin finite element method for the problem (2.1), (2.3), which is applicable on a *simplicial* mesh, say $\widetilde{\mathcal{T}}$, with respective finite element space $\widetilde{V}_h \subset V$, reading: find $\widetilde{u}_h \in \widetilde{V}_h$, such that

$$a_h\left(\widetilde{u}_h, \widetilde{v}_h\right) = \ell_h\left(\widetilde{v}_h\right), \quad \text{for all} \quad \widetilde{v}_h \in \widetilde{V}_h; \tag{4.1}$$

an example of a stable such conforming method is the streamline upwind Petrov–Galerkin approach presented and analysed in [29]. Note that the test and trial spaces for (4.1) are the same for simplicity. This can be achieved, for instance, by enforcing essential boundary conditions weakly. The abstract analysis below, however, generalizes immediately also to the case of different trial and test spaces as is often required in the treatment of non-homogeneous essential boundary conditions. Also, the space V of exact solutions is an appropriate subspace of the graph space $G := \left\{v \in L^2(\Omega) : b \cdot \nabla v \in L^2(\Omega) \quad \text{and} \quad a\nabla v \in L^2(\Omega)\right\}$.

Suppose also that the bilinear form a_h is coercive in $\widetilde{V}_h$ with respect to an "energy"-like norm $\|\cdot\|_a$, *i.e.*, for all $w \in \widetilde{V}_h$, there exists a $C_{\text{coer}} > 0$, such that

$$C_{\text{coer}} \|w\|_a^2 \leq a_h(w, w), \tag{4.2}$$

and that a is also continuous in $V \times \widetilde{V}_h$, in the sense that for all $z \in V$ and all $w \in \widetilde{V}_h$, there exists a constant $C_{\text{cont}} > 0$, such that

$$|a_h(z, w)| \leq C_{\text{cont}} |||z|||_a \|w\|_a, \tag{4.3}$$

for some norm $|||\cdot|||_a$, possibly different to $\|\cdot\|_a$. Of course, this is relevant if $|||\cdot|||_a$ is stronger than $\|\cdot\|_a$ for, otherwise, we can replace $|||\cdot|||_a$ by $\|\cdot\|_a$ throughout this section. Hence, without any loss of generality in this context, we henceforth assume $\|w\|_a \leq C_{\text{eq}}|||w|||_a$ for all $w \in V$ with $C > 0$ independent of w.

A corresponding *recovered finite element method* (R-FEM) can be, then, defined on a *polytopic* mesh $\mathcal{T}$ (with respective finite element space V_h^r), which admits a subtriangulation $\widetilde{\mathcal{T}}$ (and a respective space $\widetilde{V}_h^r$), as discussed in detail in Section 3 with the help of a recovery operator $\mathcal{E} : V_h^r \to V \cap \widetilde{V}_h^r$. To this end, we consider the R-FEM: find $u_h \in V_h^r$ (and, consequently, also $\mathcal{E}(u_h) \in V \cap \widetilde{V}_h^r$) such that

$$a_h(\mathcal{E}(u_h), \mathcal{E}(v_h)) + s_h(u_h, v_h) = \ell_h(\mathcal{E}(v_h)), \quad \text{for all} \quad v_h \in V_h^r, \tag{4.4}$$

for $s_h : \left(\widetilde{V}_h^r \cap H^1(\Omega, \mathcal{T})\right) \times \left(\widetilde{V}_h^r \cap H^1(\Omega, \mathcal{T})\right) \to \mathbb{R}$ a bilinear form, henceforth referred to as the *stabilization*, whose role is to remove the possible rank-deficiency due to the use of a recovery operator. Note that $V_h^r \subset \widetilde{V}_h^r \cap H^1(\Omega, \mathcal{T})$.

An immediate choice for stabilization can be:

$$s_h(w_h, v_h) = C \int_\Omega \mathbf{h}^m (w_h - \mathcal{E}(w_h))(v_h - \mathcal{E}(v_h)) \, \mathrm{d}x, \tag{4.5}$$

for $v_h \in V_h^r$, with $m \in \mathbb{R}$ a real number, to be determined by the error analysis in each case. When $\mathcal{E}$ is as in Lemma 3.1, (3.3) allows also to consider the alternative stabilization

$$s_h(w_h, v_h) = C \int_{\Gamma_{\text{int}}} \mathbf{h}^{m-1} [w_h] \cdot [v_h] \, \mathrm{d}s. \tag{4.6}$$

To keep the discussion general at this point, we avoid prescribing a specific stabilization, and we prefer to make a structural assumption on s_h instead.

Assumption 4.1. *The stabilization bilinear form satisfies*

$$s_h(w_h, v_h) \leq C_{\text{stab}} (s_h(w_h, w_h))^{1/2} (s_h(v_h, v_h))^{1/2} \quad \text{for all} \quad w_h, v_h \in V_h^r,$$

for some constant $C_{\text{stab}} > 0$ independent of w_h, v_h and of $\mathbf{h}$.

We also make the following stability and consistency assumptions required in the error analysis below.

Assumption 4.2.(A) *There exists a "broken" version of $\|\cdot\|_a$, say $\|\cdot\|_{a,\mathcal{T}}$, elementwise with respect to $\mathcal{T}$, for which we have $\|w\|_{a,\mathcal{T}} = \|w\|_a$ whenever $w \in \widetilde{V}_h^r \cap V$. Moreover, for $C_{\text{ker}}, c_{\text{ker}} > 0$ representing constants independent of $\mathbf{h}$ and of w the stabilization s_h satisfies*

$$c_{\text{ker}} s_h(w, w) \leq \|w - \mathcal{E}(w)\|_{a,\mathcal{T}}^2 \leq C_{\text{ker}} s_h(w, w) \quad \text{for all} \quad w \in \widetilde{V}_h^r \cap H^1(\Omega, \mathcal{T}).$$

That is, this equivalence holds for all elementwise polynomials defined over the simplicial submesh $\mathcal{T}$ that are continuous within in each element T of the related polytopic mesh.

(B) *Assume that there exists a "broken" version of $|||\cdot|||_a$, denoted by $|||\cdot|||_{a,\mathcal{T}}$, elementwise with respect to $\mathcal{T}$, for which we have:*

(1) $|||w|||_{a,\mathcal{T}} = |||w|||_a$ *whenever* $w \in \widetilde{V}_h^r \cap V$ *and*

(2) $\|w\|_{a,\mathcal{T}} \leq C|||w|||_{a,\mathcal{T}}$ *for all* $w \in \widetilde{V}_h^r \cap H^1(\Omega, \mathcal{T})$ *for some* $C > 0$ *independent of* w *and* $\mathbf{h}$.

Now given the PDE problem (2.1), (2.3) in weak form, reading: find $u \in V$ such that

$$a(u, v) = \ell(v) \quad \text{for all} \quad v \in V. \tag{4.7}$$

The following best approximation result holds.

Lemma 4.3. *Let $u \in V \cap H^1(\Omega, \mathcal{T})$ satisfy (4.7) and, for $r \geq 1$, suppose $u_h \in V_h^r$ is the R-FEM approximation with a stabilisation term satisfying Assumption 4.1 then we have*

$$\frac{1}{2}\left\|u - \mathcal{E}\left(u_h\right)\right\|_a^2 + C_{\mathrm{coer}}^{-1} s_h\left(u_h, u_h\right) \leq \inf_{v_h \in V_h^r} \left(\left(1 + 2\frac{C_{\mathrm{cont}}^2}{C_{\mathrm{coer}}^2}\right) |||u - \mathcal{E}\left(v_h\right)|||_a^2 + \frac{C_{\mathrm{stab}}^2}{C_{\mathrm{coer}}} s_h\left(v_h, v_h\right) \right) + \mathrm{INC}(u), \tag{4.8}$$

where

$$\mathrm{INC}(u) := \frac{2}{C_{\mathrm{coer}}^2} \sup_{0 \neq \widetilde{w}_h \in \mathcal{E}\left(V_h^r\right)} \left(\frac{\left|\ell_h\left(\widetilde{w}_h\right) - \ell\left(\widetilde{w}_h\right)\right|}{\|\widetilde{w}_h\|_a} + \frac{\left|a\left(u, \widetilde{w}_h\right) - a_h\left(u, \widetilde{w}_h\right)\right|}{\|\widetilde{w}_h\|_a} \right)^2,$$

and $C_{\mathrm{coer}}, C_{\mathrm{cont}}$ are the coercivity and continuity constants defined in (4.2), (4.3) independent of u, u_h, h and of $\mathcal{E}$.

Proof. With $\xi := \mathcal{E}\left(v_h - u_h\right) \in \widetilde{V}_h^r \cap V =: \widetilde{V}_h$, coercivity (4.2) implies

$$C_{\mathrm{coer}} \|\xi\|_a^2 + s_h\left(u_h, u_h\right) \leq a_h(\xi, \xi) + s_h\left(u_h, u_h\right),$$

and hence, in view of (4.4), (4.7) adding and subtracting appropriate terms yields

$$\begin{aligned} C_{\mathrm{coer}} \|\xi\|_a^2 + s_h\left(u_h, u_h\right) &\leq a_h(\mathcal{E}\left(v_h\right), \xi) - \ell_h(\xi) + s_h\left(u_h, v_h\right) \\ &\leq a_h(\mathcal{E}\left(v_h\right) - u, \xi) + s_h\left(u_h, v_h\right) + \ell(\xi) - \ell_h(\xi) + a_h(u, \xi) - a(u, \xi). \end{aligned}$$

Making use of the continuity (4.3) of a_h along with Assumption 4.1 we see

$$\begin{aligned} \|\xi\|_a^2 + \frac{1}{C_{\mathrm{coer}}} s_h\left(u_h, u_h\right) \leq \frac{C_{\mathrm{cont}}}{C_{\mathrm{coer}}} |||u - \mathcal{E}\left(v_h\right)|||_a \|\xi\|_a + \frac{C_{\mathrm{stab}}}{C_{\mathrm{coer}}} \left(s_h\left(u_h, u_h\right)\right)^{1/2} \left(s_h\left(v_h, v_h\right)\right)^{1/2} \\ + \frac{1}{C_{\mathrm{coer}}} \left(\ell(\xi) - \ell_h(\xi) + a_h(u, \xi) - a(u, \xi)\right). \end{aligned}$$

Finally, invoking Hölder's inequality in standard fashion shows the abstract bound

$$\|\xi\|_a^2 + \frac{1}{C_{\mathrm{coer}}} s_h\left(u_h, u_h\right) \leq 2\frac{C_{\mathrm{cont}}^2}{C_{\mathrm{coer}}^2} |||u - \mathcal{E}\left(v_h\right)|||_a^2 + \frac{C_{\mathrm{stab}}^2}{C_{\mathrm{coer}}} s_h\left(v_h, v_h\right) + \mathrm{INC}(u).$$

The result follows by the triangle inequality and noticing that v_h was arbitrary, together with $\|w\|_a \leq C_{\mathrm{eq}} |||w|||_a$ for all $w \in V$. □

To arrive at an *a priori* error bound, we make the following (rather mild and immediately satisfiable by all the scenarios we have in mind) additional set of assumptions.

Theorem 4.4. *Assume that the recovery operator $\mathcal{E}$ in the definition of R-FEM (4.4) is such that $\mathcal{E}(v) = v$ for all $v \in \widetilde{V}_h^r \cap V$ and, also, that it is stable with respect to the $|||w|||_{a,\mathcal{T}}$-norm, viz.,*

$$|||\mathcal{E}(w)|||_a \leq C |||w|||_{a,\mathcal{T}} \quad \forall w \in V_h^r.$$

Assume that the exact solution satisfies $u|_T \in H^k(T), T \in \mathcal{T}$, for some $k \geq 2$, and that any inconsistency of the Galerkin method posed on simplices (4.1) is of optimal order, viz.,

$$\mathrm{INC}(u) \leq C \sum_{T \in \widetilde{\mathcal{T}}} \mathbf{h}^{2s-\alpha} |u|_{s,T}^2,$$

for $1 \leq s = \min\{k, r+1\}$, $\alpha \in \{1, 2\}$ depending on the structure of the diffusion tensor a, with constant $C > 0$, independent of u and $\mathbf{h}$. Then, we have the bound

$$\|u - \mathcal{E}\left(u_h\right)\|_a^2 + s_h\left(u_h, u_h\right) \leq C \sum_{T \in \widetilde{\mathcal{T}}} \mathbf{h}^{2s-\alpha} |u|_{s,T}^2. \tag{4.9}$$

Proof. The triangle inequality, Assumption 4.2 along with the optimality of the inconsistency terms imply

$$\|u - \mathcal{E}(u_h)\|_a^2 + s_h(u_h, u_h) \leq C|||u - \Pi u|||_{a,\mathcal{T}}^2 + Cs_h(\Pi u, \Pi u) + C\sum_{T\in\widetilde{\mathcal{T}}} \mathbf{h}^{2s-\alpha}|u|_{s,T}^2, \tag{4.10}$$

with $\Pi : L^2(\Omega) \to V_h^r$ denoting the orthogonal L^2-projection operator onto the polytopic finite element space V_h^r. Let also $\widetilde{\Pi} : L^2(\Omega) \to \widetilde{V}_h^r \cap V$ be the respective orthogonal L^2-projection onto the conforming finite element space of the subtriangulation $\widetilde{\mathcal{T}}$. The above mesh assumptions on $\mathcal{T}$ and on $\widetilde{\mathcal{T}}$ ensure that Π and $\widetilde{\Pi}$ admit optimal approximation properties.

From hypothesis, we have $\mathcal{E}(\widetilde{\Pi}u) = \widetilde{\Pi}u$. From Assumption 4.2, it then follows $s_h\left(\widetilde{\Pi}u, \widetilde{\Pi}u\right) = 0$. Now, from Assumption 4.1, we then also have

$$s_h\left(\widetilde{\Pi}u, v\right) = s_h\left(v, \widetilde{\Pi}u\right) = 0 \quad \text{for any} \quad v \in \widetilde{V}_h^r \cap H^1(\Omega, \mathcal{T}).$$

Using this, together with Assumption 4.2 and the stability of the recovery operator, we have, respectively,

$$\begin{aligned} s_h(\Pi u, \Pi u) &= s_h\left(\widetilde{\Pi}u - \Pi u, \widetilde{\Pi}u - \Pi u\right) \\ &\leq c_{\text{ker}}^{-1}\left\|\widetilde{\Pi}u - \Pi u - \mathcal{E}\left(\widetilde{\Pi}u - \Pi u\right)\right\|_{a,\mathcal{T}}^2 \\ &\leq C\left(\left\|\widetilde{\Pi}u - \Pi u\right\|_{a,\mathcal{T}}^2 + |||\widetilde{\Pi}u - \Pi u|||_{a,\mathcal{T}}^2\right) \\ &\leq C|||\widetilde{\Pi}u - \Pi u|||_{a,\mathcal{T}}^2 \\ &\leq C\left(|||u - \widetilde{\Pi}u|||_{a,\mathcal{T}} + |||u - \Pi u|||_{a,\mathcal{T}}\right)^2. \end{aligned}$$

The result now follows by appealing to the optimal approximation properties of Π and $\widetilde{\Pi}$. □

The nature of the diffusion tensor a determines the strength of the norms $\|\cdot\|_a$ and $|||\cdot|||_a$. For instance, for the case of the streamline upwinded Petrov Galerkin method of [29], we can take

$$\|v\|_a = |||v|||_a = \left(\gamma_0\|v\|_\Omega^2 + \left\|\sqrt{\tau}\mathcal{L}v\right\|_\Omega^2 + \left\|\sqrt{b\cdot n}u\right\|_{\Gamma_-\cup\Gamma_+}^2\right)^{1/2},$$

and

$$|||v|||_{a,\mathcal{T}} = \left(\gamma_0\|v\|_\Omega^2 + \sum_{T\in\mathcal{T}}\left\|\sqrt{\tau}\mathcal{L}v\right\|_T^2 + \left\|\sqrt{b\cdot n}u\right\|_{\Gamma_-\cup\Gamma_+}^2\right)^{1/2},$$

with $\tau|_T = ch_T/p$, for some appropriate $c > 0$. When no diffusion is present ($a \equiv 0$), Theorem 4.4 holds with $\alpha = 1$, while when $a \neq 0$, optimal rates are still given for $\alpha = 2$.

Corollary 4.5. *With the assumptions of Theorem 4.4, we also have the following bound:*

$$\|u - u_h\|_{a,\mathcal{T}}^2 + s_h(u_h, u_h) \leq C\sum_{T\in\widetilde{\mathcal{T}}} \mathbf{h}^{2s-\alpha}|u|_{s,T}^2, \tag{4.11}$$

for $1 \leq s = \min\{k, r\}$, *with* C *positive constant, independent of* u *and of* $\mathbf{h}$.

Proof. The triangle inequality implies

$$\|u - u_h\|_{a,\mathcal{T}} \leq \|u - \mathcal{E}(u_h)\|_a + \|\mathcal{E}(u_h) - u_h\|_{a,\mathcal{T}}.$$

Using, now, Assumption 4.2 we have

$$\|\mathcal{E}(u_h) - u_h\|_{a,\mathcal{T}} \leq C_{\text{ker}}s_h(u_h, u_h),$$

the result then follows. □

5. An alternative recovered finite element method

To highlight further the potential of the proposed R-FEM framework applied to both standard/simplicial/box and, in general, polytopic meshes, we present an alternative R-FEM. This method is motivated by the desire to have a conforming approximation for the second order part of the differential operator and an upwinded discontinuous Galerkin discretization of the first order terms in (2.1). The developments below also showcase an R-FEM error analysis using inf-sup stability rather than coercivity results.

For sake of the simplicity of exposition, we assume that each entries of the diffusion tensor a are constant on each element $T \in \mathcal{T}$,*i.e.*,

$$a \in \left[V_h^0\right]_{\text{sym}}^{d\times d}. \tag{5.1}$$

Additionally, we assume the following standard assumption on $\mathbf{b}$:

$$\mathbf{b}\cdot\nabla_h\xi \in V_h^r \quad \forall \xi \in V_h^r. \tag{5.2}$$

cf., [11,31]. For given operator (3.2), the new *recovered finite element method* reads: find $u_h \in V_h^r$ such that

$$B\left(u_h, v_h\right) = \ell\left(v_h\right), \quad \text{for all} \quad v_h \in V_h^r, \tag{5.3}$$

where

$$\begin{aligned} B\left(u_h, v_h\right) := & \int_\Omega \left(a\nabla\mathcal{E}\left(u_h\right)\cdot\nabla\mathcal{E}\left(v_h\right) + \mathbf{b}\cdot\nabla_h u_h\mathcal{E}\left(v_h\right) + c\mathcal{E}\left(u_h\right)\mathcal{E}\left(v_h\right)\right)\mathrm{d}x \\ & - \int_{\Gamma_{\mathrm{D}}} \left(a\nabla\mathcal{E}\left(u_h\right)\cdot\mathbf{n}\,\mathcal{E}\left(v_h\right) + a\nabla\mathcal{E}\left(v_h\right)\cdot\mathbf{n}\,\mathcal{E}\left(u_h\right) - \sigma_{\mathrm{D}}\mathcal{E}\left(u_h\right)\mathcal{E}\left(v_h\right)\right)\mathrm{d}s \\ & - \int_{\Gamma_{\mathrm{D}}^-} \left(\mathbf{b}\cdot\mathbf{n}\right)u_h\mathcal{E}\left(v_h\right)\mathrm{d}s - \sum_{T\in\mathcal{T}}\int_{\partial_- T\setminus\partial\Omega} \left(\mathbf{b}\cdot\mathbf{n}\right)\lfloor u_h\rfloor\mathcal{E}\left(v_h\right)\mathrm{d}s \\ & + s_h^{a,c}\left(u_h, v_h\right) + s_h^b\left(u_h, v_h\right), \end{aligned} \tag{5.4}$$

and

$$\begin{aligned} \ell\left(v_h\right) := & \int_\Omega f\mathcal{E}\left(v_h\right)\,\mathrm{d}x - \int_{\Gamma_{\mathrm{D}}} g_{\mathrm{D}}\left(a\nabla\mathcal{E}\left(v_h\right)\cdot\mathbf{n} - \sigma_{\mathrm{D}}\mathcal{E}\left(v_h\right)\right)\mathrm{d}s + \int_{\Gamma_{\mathrm{N}}} g_{\mathrm{N}}\,\mathcal{E}\left(v_h\right)\mathrm{d}s \\ & - \int_{\Gamma_{\mathrm{D}}^-} \left(\mathbf{b}\cdot\mathbf{n}\right)g_{\mathrm{D}}\mathcal{E}\left(v_h\right)\,\mathrm{d}s, \end{aligned}$$

with $s_h^m(\cdot,\cdot) : V_h^r \times V_h^r \to \mathbb{R}$, $m \in \{\{a,c\}, b\}$ denoting symmetric bilinear forms, henceforth referred to as *stabilisations*, and $\sigma_{\mathrm{D}} : \Gamma_{\mathrm{D}} \to \mathbb{R}$ a positive penalty function defined precisely below that weakly enforces the Dirichlet boundary conditions.

This motivates the following choice for the elliptic stabilisation bilinear form:

$$s_h^{a,c}\left(u_h, v_h\right) := \int_{\Gamma_{\text{int}}} \sigma_{a,c}\left[u_h\right]\cdot\left[v_h\right]\,\mathrm{d}s, \tag{5.5}$$

for some non-negative function $\sigma_{a,c} : \Gamma_{\text{int}} \to \mathbb{R}$, that will also be defined below.

To ensure that sufficient numerical diffusion is included in the proposed method for the case of small or vanishing diffusion tensor a, we select

$$s_h^b\left(u_h, v_h\right) := \int_{\Gamma_{\text{int}}} \left(\sigma_{b,1}\left[u_h\right]\cdot\left[v_h\right] + \sigma_{b,2}\left[\mathbf{h}\left(\mathbf{b}\cdot\nabla u_h\right)\right]\cdot\left[\mathbf{h}\left(\mathbf{b}\cdot\nabla v_h\right)\right]\right)\mathrm{d}s, \tag{5.6}$$

for non-negative functions $\sigma_{b,1}, \sigma_{b,2} : \Gamma_{\text{int}} \to \mathbb{R}$, to be selected below. We note that (5.6) follows the spirit of the, so-called, continuous interior penalty stabilisation procedure due to Douglas and Dupont [22] and to Burman

and Hansbo [9]. Crucially, however, the trial and test functions u_h and v_h in the present R-FEM context are discontinuous, *cf.*, [8] also. The inconsistency introduced by the streamline derivative jump term in (5.6) will be dealt with in the *a priori* error analysis below.

Some remarks on the method are in order. To accommodate for the potentially locally changing nature of the differential operator, we have opted for weak imposition of essential boundary conditions, following the classical ideas from [34,38]. For the case of elliptic problems, strong imposition of essential boundary conditions in the spirit of [24] is by all means possible also. We note, however, that since essential boundary values are known the above is actually a conforming method for $\mathcal{E}(u_h)$.

Also, we have opted for a, to the best of our knowledge, new method for the discretisation of the first order term. This is to highlight the flexibility of R-FEM in incorporating different discretisations of various terms of the differential operator. More importantly, since in the absence of diffusion, the exact solution u may exhibit jump discontinuities across characteristic surfaces, we prefer not to recover u_h in the discretisation of the first order term. We stress that more classical choices, such as streamline diffusion-type and/or continuous interior penalty-type treatment of the first order term are by all means possible. Indeed, certain such choices coincide with the standard/classical conforming finite element versions for $\mathcal{E}(u_h)$ when applied to standard triangular meshes (*cf.*, the discussion in Section 3 of [24]).

We also remark on the assumptions on the diffusion tensor (5.1) and convection field (5.2). The above R-FEM method (5.3) can be easily extended to general positive semi-definite diffusion $a \in [L^\infty(\Omega)]^{d\times d}_{\text{sym}}$ following the inconsistent formulation introduced in [23]. For the general convection field $\mathbf{b}$, we would need to modify (5.6) by setting

$$s_h^b(u_h, v_h) = \int_{\Gamma_{\text{int}}} (\sigma_{b,1}[u_h]\cdot[v_h] + \sigma_{b,2}[\mathbf{h}\Pi(\mathbf{b}\cdot\nabla u_h)]\cdot[\mathbf{h}\Pi(\mathbf{b}\cdot\nabla v_h)])\,\mathrm{d}s, \tag{5.7}$$

which will make the stability proof and error analysis more complicated. We refrain from doing this here to focus on the key ideas.

Finally, some comments on the practicality of the R-FEM framework for polytopic meshes, compared to other approaches is in order. The total number of unknowns for the R-FEM depends only on the number of polygonal elements and on the local polynomial degree, but is independent of the number of faces per element. Therefore, R-FEM may be advantageous in when elements with many faces are present in the mesh and, thus, be a "conforming" alternative to interior penalty discontinuous Galerkin methods (IP-dG) for many-face-per-element polytopic meshes [12]; see also the numerical experiment in Section 7.5.

At the same time, R-FEM admits a wider stencil than an IP-dG method on the same mesh. To counteract this, it would certainly be an interesting direction of future research to derive hybridized versions of R-FEM in the spirit of [19,21]. Nonetheless, R-FEM is primarily envisaged to eventually be able to tackle problems requiring control of compatibility properties of the approximated state variable (such as positivity, monotonicity and conservation properties) on meshes which result from simplicial mesh agglomeration, whereby the individual element shapes are difficult to control. To that end, the proof of R-FEM stability under face and/or edge degeneration is a very interesting direction of future research.

6. *A PRIORI* ERROR ANALYSIS

We dedicate this section to the analysis of the method introduced in Section 5. The main ingredient to this is an inf-sup condition over suitable norms. We let $\alpha, \beta, \gamma : \Omega \to \mathbb{R}$ such that

$$\alpha|_T := |\sqrt{a}|_2^2|_T, \quad \beta|_T := \|\mathbf{b}\|_{L^\infty(T)}, \quad \gamma|_T := \|c\|_{L^\infty(T)}, \tag{6.1}$$

over each element $T \in \mathcal{T}$. We define the stabilisation parameter

$$\sigma_{\text{D}} := C_\sigma \alpha r^2/\mathbf{h}, \quad \widetilde{\sigma}_{\text{D}}|_T := \max_{e\subset\partial T} \sigma_{\text{D}}|_e. \tag{6.2}$$

Now, for $w \in V_h^r$, we define the norms

$$\|w\|_{\mathrm{b}} := \left(\|\sqrt{c_0} w\|^2 + \frac{1}{2} \left(\left\| \sqrt{|\mathbf{b}\cdot\mathbf{n}|} \lfloor w \rfloor \right\|_{\Gamma_{\mathrm{int}}}^2 + \left\| \sqrt{|\mathbf{b}\cdot\mathbf{n}|} w \right\|_{\Gamma_{\mathrm{D}}^-}^2 + \left\| \sqrt{|\mathbf{b}\cdot\mathbf{n}|} w \right\|_{\Gamma_{\mathrm{N}}^+}^2 \right) \right)^{1/2},$$

and

$$|||w||| := \left(\left\| \sqrt{a} \nabla \mathcal{E}(w) \right\|^2 + \|\sqrt{\sigma_{\mathrm{D}}} \mathcal{E}(w)\|_{\Gamma_{\mathrm{D}}}^2 + \|w\|_{\mathrm{b}}^2 + s_h^{a,c}(w,w) + s_h^b(w,w) \right)^{1/2}.$$

We also define the "streamline-diffusion" norm

$$|||w|||_s := \left(|||w|||^2 + \left\| \sqrt{\delta \boldsymbol{\lambda}} \left(\mathbf{b} \cdot \nabla_h w \right) \right\|^2 \right)^{1/2},$$

where $\delta > 0$, to be chosen precisely below, and

$$\boldsymbol{\lambda} := \min \left\{ \beta^{-1}, \widetilde{\sigma}_{\mathrm{D}}^{-1} \right\} \mathbf{h}, \tag{6.3}$$

We are now in a position to show the following inf-sup condition for the R-FEM method (5.3). In the proofs of the results in this section, we are particularly interested in the dependence of the resulting bounds on the mesh-Péclet number Pe_h, the mesh-size $\mathbf{h}$ and polynomial degree r. We aim, therefore, to track constants and their dependence explicitly.

Theorem 6.1 (Inf-Sup Condition)**.** *Let $a \in [L^\infty(\Omega)]_{\mathrm{sym}}^{d\times d}$ satisfy assumption (5.1) and $\mathbf{b} \in W^{1,\infty}(\Omega)^d$ satisfy assumption (5.2). Assume that the mesh is such that each element face in the mesh is either completely inflow or outflow or characteristic. Suppose also that the penalisation parameters $\sigma_{a,c}, \sigma_{b,1}$ and $\sigma_{b,2}$ are chosen large enough to satisfy (6.10), δ is chosen to satisfy (6.11) and the boundary stabilisation constant $C_\sigma > 0$ is sufficiently large. Then, we have*

$$\inf_{0\neq w_h \in V_h^r} \sup_{0\neq v_h \in V_h^r} \frac{B(w_h, v_h)}{|||w_h|||_s |||v_h|||_s} \geq \Lambda, \tag{6.4}$$

where $\Lambda > 0$ is independent of $\boldsymbol{\lambda}$, $\mathbf{h}$ and of the mesh-Péclet number $\mathrm{Pe}_h := \beta\mathbf{h}/\alpha$.

Proof. As usual, the proof consists of two steps: (1) for each $w_h \in V_h^r$, we find a $v_h(w_h) \equiv v_h \in V_h^r$ such that $B(w_h, v_h) \geq C|||w_h|||_s^2$, and, (2) we show that this v_h satisfies the bound $|||v_h|||_s \leq C|||w_h|||_s$, thereby inferring the result.

To that end, fix $w_h \in V_h^r$ and set $v_h = w_h + \delta w_h^b$, where we will use the shorthand $w_h^b := \boldsymbol{\lambda}(\mathbf{b}\cdot\nabla_h w_h)$ for brevity for some $\delta \in \mathbb{R}$ is to be chosen. Then, integration by parts and working as in the proof of Lemma 2.4 from [30], as well as making use of standard inverse estimates, give

$$\begin{aligned} B(w_h, w_h) &\geq \frac{1}{2}|||w_h|||^2 + \int_\Omega \mathbf{b}\cdot\nabla_h w_h \left(\mathcal{E}(w_h) - w_h \right) \mathrm{d}x + \int_\Omega c \left(\mathcal{E}(w_h)^2 - w_h^2 \right) \mathrm{d}x \\ &\quad - \int_{\Gamma_{\mathrm{D}}^-} (\mathbf{b}\cdot\mathbf{n}) w_h \left(\mathcal{E}(w_h) - w_h \right) \mathrm{d}s - \sum_{T\in\mathcal{T}} \int_{\partial_- T \setminus \partial\Omega} (\mathbf{b}\cdot\mathbf{n}) \lfloor w_h \rfloor \left(\mathcal{E}(w_h) - w_h \right) \mathrm{d}s \\ &=: \frac{1}{2}|||w_h|||^2 + \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV}. \end{aligned} \tag{6.5}$$

Using Lemma 3.1, Young's inequality and (6.3), we have

$$\mathrm{I} \leq \frac{1}{4} \left\| \sqrt{\delta\boldsymbol{\lambda}} \mathbf{b} \cdot \nabla_h w_h \right\|^2 + C(r) \left\| \delta^{-1/2} \beta^{1/2} [w_h] \right\|_{\Gamma_{\mathrm{int}}}^2.$$

and

$$\begin{aligned}\mathrm{II} &= \int_\Omega c\left(\mathcal{E}\left(w_h\right)-w_h\right)^2+2cw_h\left(\mathcal{E}\left(w_h\right)-w_h\right)\mathrm{d}x\\ &\le C(r)\left\|\sqrt{\gamma\mathbf{h}}\left[w_h\right]\right\|_{\Gamma_{\mathrm{int}}}^2+\frac{1}{4}\left\|\sqrt{c_0}w_h\right\|^2+C(r)\left\|\sqrt{\gamma/\underline{\gamma}_0}\sqrt{\gamma\mathbf{h}}\left[w_h\right]\right\|_{\Gamma_{\mathrm{int}}}^2\\ &\le \frac{1}{4}\left\|\sqrt{c_0}w_h\right\|^2+C(r)\left\|\left(1+\sqrt{\gamma/\underline{\gamma}_0}\right)\sqrt{\gamma\mathbf{h}}\left[w_h\right]\right\|_{\Gamma_{\mathrm{int}}}^2,\end{aligned}$$

with $\underline{\gamma}_0$ a local min of c_0. Finally, again, using Lemma 3.1 and Young's inequality, we also have

$$\mathrm{III}+\mathrm{IV}\le\frac{1}{8}\left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}w_h\right\|_{\Gamma_{\mathrm{D}}^-}^2+\frac{1}{8}\left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}\lfloor w_h\rfloor\right\|_{\Gamma_{\mathrm{int}}}^2+C(r)\left\|\sqrt{\beta}\left[w_h\right]\right\|_{\Gamma_{\mathrm{int}}}^2.$$

Substituting the above bounds into (6.5), we arrive at

$$\begin{aligned}B\left(w_h,w_h\right)\ge{}&\frac{1}{2}|||w_h|||^2-\frac{1}{4}\left\|\sqrt{c_0}w_h\right\|^2-C(r)\left\|\left(1+\sqrt{\gamma/\underline{\gamma}_0}\right)\sqrt{\gamma\mathbf{h}}\left[w_h\right]\right\|_{\Gamma_{\mathrm{int}}}^2\\ &-\frac{1}{8}\left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}w_h\right\|_{\Gamma_{\mathrm{D}}^-}^2-\frac{1}{8}\left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}\lfloor w_h\rfloor\right\|_{\Gamma_{\mathrm{int}}}^2-\frac{1}{4}\left\|\sqrt{\delta\boldsymbol{\lambda}}\mathbf{b}\cdot\nabla_h w_h\right\|^2\\ &-C(r)\left\|\left(\sqrt{\beta}+\delta^{-1/2}\sqrt{\beta}\right)\left[w_h\right]\right\|_{\Gamma_{\mathrm{int}}}^2.\end{aligned}\tag{6.6}$$

Working as before, we also have

$$\begin{aligned}B\left(w_h,\delta w_h^b\right)\ge{}&-\frac{1}{4}|||w_h|||-\left\|\sqrt{a}\nabla\mathcal{E}\left(\delta w_h^b\right)\right\|^2\\ &+\left\|\sqrt{\delta\boldsymbol{\lambda}}\mathbf{b}\cdot\nabla_h w_h\right\|^2+\int_\Omega\mathbf{b}\cdot\nabla_h w_h\left(\mathcal{E}\left(\delta w_h^b\right)-\delta w_h^b\right)\mathrm{d}x\\ &-2\left\|\sqrt{\sigma_{\mathrm{D}}}\mathcal{E}\left(\delta w_h^b\right)\right\|_{\Gamma_{\mathrm{D}}}^2-8\left\|\sigma_{\mathrm{D}}^{-1/2}a\nabla\mathcal{E}\left(\delta w_h^b\right)\right\|_{\Gamma_{\mathrm{D}}}^2\\ &-2\left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}\mathcal{E}\left(\delta w_h^b\right)\right\|_{\Gamma_{\mathrm{D}}^-}^2-2\left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}\mathcal{E}\left(\delta w_h^b\right)\right\|_{\Gamma_{\mathrm{int}}}^2\\ &-\left\|\sqrt{\sigma_{a,c}}\left[\delta w_h^b\right]\right\|_{\Gamma_{\mathrm{int}}}^2-\left\|\sqrt{\sigma_{b,1}}\left[\delta w_h^b\right]\right\|_{\Gamma_{\mathrm{int}}}^2-\left\|\sqrt{\sigma_{b,2}}\left[\mathbf{h}\left(\mathbf{b}\cdot\nabla\left(\delta w_h^b\right)\right)\right]\right\|_{\Gamma_{\mathrm{int}}}^2.\end{aligned}\tag{6.7}$$

We further bound each term in (6.7) not directly appearing in the energy norm. We have

$$\begin{aligned}\left\|\sqrt{a}\nabla\mathcal{E}\left(\delta w_h^b\right)\right\|^2&\le Cr^4\left\|\sqrt{\alpha}\mathbf{h}^{-1}\mathcal{E}\left(\delta w_h^b\right)\right\|^2\\ &\le Cr^4\left\|\sqrt{\delta}\left(\sqrt{\delta\boldsymbol{\lambda}}\mathbf{b}\cdot\nabla_h w_h\right)\right\|^2+C(r)\left\|\left[\delta\mathbf{h}\left(\mathbf{b}\cdot\nabla_h w_h\right)\right]\right\|_{\Gamma_{\mathrm{int}}}^2,\end{aligned}\tag{6.8}$$

using an inverse estimate and Lemma 3.1, respectively. Similarly,

$$\int_\Omega\mathbf{b}\cdot\nabla_h w_h\left(\mathcal{E}\left(\delta w_h^b\right)-\delta w_h^b\right)\mathrm{d}x\le\frac{1}{4}\left\|\sqrt{\delta\boldsymbol{\lambda}}\mathbf{b}\cdot\nabla_h w_h\right\|^2+C(r)\left\|\left[\sqrt{\delta}\mathbf{h}\left(\mathbf{b}\cdot\nabla_h w_h\right)\right]\right\|_{\Gamma_{\mathrm{int}}}^2.$$

Next, using the stability of $\mathcal{E}$, we deduce

$$2\left\|\sqrt{\sigma_{\mathrm{D}}}\mathcal{E}\left(\delta w_h^b\right)\right\|_{\Gamma_{\mathrm{D}}}^2\le Cr^4\left\|\sqrt{\delta}\left(\sqrt{\delta\boldsymbol{\lambda}}\mathbf{b}\cdot\nabla_h w_h\right)\right\|^2+C(r)\left\|\left[\delta\mathbf{h}\left(\mathbf{b}\cdot\nabla_h w_h\right)\right]\right\|_{\Gamma_{\mathrm{int}}}^2,$$

and, using inverse estimates along with the definition of σ_{D},

$$8\left\|\sigma_{\mathrm{D}}^{-1/2}a\nabla\mathcal{E}\left(\delta w_h^b\right)\right\|_{\Gamma_{\mathrm{D}}}^2 \leq Cr^4\left\|\sqrt{\alpha}\mathbf{h}^{-1}\mathcal{E}\left(\delta w_h^b\right)\right\|^2,$$

which can be further estimated as in (6.8). We also have

$$2\left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}\mathcal{E}\left(\delta w_h^b\right)\right\|_{\Gamma_{\mathrm{D}}^-\cup\Gamma_{\mathrm{int}}}^2 \leq Cr^2\left\|\sqrt{\delta}\left(\sqrt{\delta\boldsymbol{\lambda}}\mathbf{b}\cdot\nabla_h w_h\right)\right\|^2 + C(r)\left\|[\delta\mathbf{h}\left(\mathbf{b}\cdot\nabla_h w_h\right)]\right\|_{\Gamma_{\mathrm{int}}}^2,$$

and

$$\left\|\sqrt{\sigma_{b,2}}\left[\mathbf{h}\left(\mathbf{b}\cdot\nabla_h\left(\delta w_h^b\right)\right)\right]\right\|_{\Gamma_{\mathrm{int}}}^2 \leq Cr^6\left\|\sqrt{\sigma_{b,2}/\mathbf{h}}\beta\delta w_h^b\right\|^2 \leq Cr^6\left\|\sqrt{\sigma_{b,2}\beta\delta}\left(\sqrt{\delta\boldsymbol{\lambda}}\mathbf{b}\cdot\nabla_h w_h\right)\right\|^2.$$

Combining now the above estimates, we arrive at the bound

$$\begin{aligned}
B\left(w_h, v_h\left(w_h\right)\right) \geq\ & \frac{1}{4}|||w_h|||^2 + \frac{1}{2}\left\|\sqrt{\delta\mathbf{h}}\mathbf{b}\cdot\nabla_h w_h\right\|^2 \\
& - C(r)\left\|\left(\sqrt{\beta}+\sqrt{\beta}\delta^{-1/2}\right)[w_h]\right\|_{\Gamma_{\mathrm{int}}}^2 - C(r)\left\|\left(1+\sqrt{\gamma/\underline{\gamma}_0}\right)\sqrt{\gamma\mathbf{h}}[w_h]\right\|_{\Gamma_{\mathrm{int}}}^2 \\
& - Cr^4\left\|\sqrt{\delta}\left(1+r^{-1}+r\sqrt{\sigma_{b,2}\beta}\right)\left(\sqrt{\delta\mathbf{h}}\mathbf{b}\cdot\nabla_h w_h\right)\right\|^2 \\
& - C(r)\left\|\left(\sqrt{\sigma_{b,1}}+1+\delta^{-1/2}\right)[\delta\mathbf{h}\left(\mathbf{b}\cdot\nabla_h w_h\right)]\right\|_{\Gamma_{\mathrm{int}}}^2.
\end{aligned} \tag{6.9}$$

Upon selecting a global constant $\delta > 0$, small enough, we can have

$$Cr^4\delta\left(1+r^{-1}+r\sqrt{\sigma_{b,2}\beta}\right)^2 \leq \frac{1}{4} \quad \text{and} \quad \delta^2 C(r)\left(\sqrt{\sigma_{b,1}}+1+\delta^{-1/2}\right)^2 \leq \frac{\sigma_{b,2}}{8}.$$

Now, upon selecting additionally penalty parameters large enough to satisfy

$$\sigma_{b,1} \geq 8C(r)\beta\left(1+\delta^{-1/2}\right)^2, \quad \sigma_{a,c} \geq 8C(r)\gamma\left(1+\sqrt{\gamma/\underline{\gamma}_0}\right)^2, \quad \sigma_{b,2} > 0, \tag{6.10}$$

we can set

$$\delta := \min\left\{\left(4Cr^4\left(1+r^{-1}+r\sqrt{\sigma_{b,2}\beta}\right)^2\right)^{-1}, \left(-1+\sqrt{1+\sqrt{\sigma_{b,2}}\left(2\sqrt{C(r)\beta}+2^{-1/2}\right)^{-1}}\right)^2/4\right\}. \tag{6.11}$$

which, in turn implies,

$$B\left(w_h, v_h\left(w_h\right)\right) \geq \frac{1}{8}|||w_h|||^2 + \frac{1}{4}\left\|\sqrt{\delta\mathbf{h}}\left(\mathbf{b}\cdot\nabla_h w_h\right)\right\|^2 \geq \frac{1}{8}|||w_h|||_s^2, \tag{6.12}$$

from (6.9) and the first step of the proof is complete.

For the second step, working as above, standard inverse estimates along with and Lemma 3.1 imply

$$\begin{aligned}
|||\delta w_h^b|||^2 \leq\ & C(r)\left\|[\delta\mathbf{h}\left(\mathbf{b}\cdot\nabla_h w_h\right)]\right\|_{\Gamma_{\mathrm{int}}}^2 + Cr^4\left\|\sqrt{\delta}\left(\sqrt{\delta\boldsymbol{\lambda}}\mathbf{b}\cdot\nabla_h w_h\right)\right\|^2 + \left\|\sqrt{c_0\delta\boldsymbol{\lambda}}\sqrt{\delta\boldsymbol{\lambda}}\mathbf{b}\cdot\nabla_h w_h\right\|^2 \\
& + \left\|\sqrt{\sigma_{a,c}}\left[\delta w_h^b\right]\right\|_{\Gamma_{\mathrm{int}}}^2 + Cr^2\left\|\sqrt{\beta\delta}\sqrt{\delta\boldsymbol{\lambda}}\mathbf{b}\cdot\nabla_h w_h\right\|_{\Gamma_{\mathrm{int}}}^2 + \left\|\sqrt{\sigma_{b,1}}\left[\delta w_h^b\right]\right\|_{\Gamma_{\mathrm{int}}}^2 \\
& + Cr^6\left\|\sqrt{\sigma_{b,2}\delta}\beta\sqrt{\delta\boldsymbol{\lambda}}\mathbf{b}\cdot\nabla_h w_h\right\|^2
\end{aligned}$$

and $\left\|\sqrt{\delta\boldsymbol{\lambda}}\mathbf{b}\cdot\nabla_h\delta w_h^b\right\|^2 \leq Cr^4\left\|\delta\sqrt{\delta\boldsymbol{\lambda}}\mathbf{b}\cdot\nabla_h w_h\right\|^2$. The above assumptions on δ, $\sigma_{a,c}$, $\sigma_{b,1}$, $\sigma_{b,2}$, finally imply

$$|||v_h|||_s \leq |||w_h|||_s + |||\delta w_h^b|||_s \leq C(r)\left(|||w_h||| + \left\|\sqrt{\delta\mathbf{h}}\mathbf{b}\cdot\nabla_h w_h\right\|\right) = C(r)|||w_h|||_s, \tag{6.13}$$

thereby completing the proof of the second step also. □

Remark 6.2. It is possible to modify the above proof by introducing a locally variable δ aiming to achieve stronger streamline-diffusion stabilization effect at the expense of mild assumptions on the local variation of δ in the computational domain, see [26] for a similar argument in a completely different context. In particular, using the elementary identity $\left[\delta w_h^b\right] = [\delta]\left\{w_h^b\right\} + \{\delta\}\left[w_h^b\right]$ valid on every face $e \subset \Gamma_{\text{int}}$, provided that $\|[\delta]\|_{L^\infty(e)} / \|\delta\|_{L^\infty(e)} \ll 1$, one can incorporate $\left\{w_h^b\right\}$ into the stabilization term $\left\|\sqrt{\delta\boldsymbol{\lambda}}\mathbf{b}\cdot\nabla_h w_h\right\|$.

Proposition 6.3 (Galerkin orthogonality)**.** *Let $u \in V$ be the solution of (2.1), (2.3). Suppose also that $u_h \in V_h^r$ is the R-FEM solution of (5.3) and set $e := u - \mathcal{E}(u_h)$ for brevity. Then, for all $v_h \in V_h^r$ we have:*

$$\begin{aligned}
&\int_\Omega \left(a\nabla e\cdot\nabla\mathcal{E}(v_h) + \mathbf{b}\cdot\nabla_h(u-u_h)\,\mathcal{E}(v_h) + ce\mathcal{E}(v_h)\right)\mathrm{d}x \\
&\quad - \int_{\Gamma_{\mathrm{D}}} \left(a\nabla e\cdot\mathbf{n}\,\mathcal{E}(v_h) + a\nabla\mathcal{E}(v_h)\cdot\mathbf{n}\,e - \sigma_{\mathrm{D}} e\mathcal{E}(v_h)\right)\mathrm{d}s \\
&\quad - \int_{\Gamma_{\mathrm{D}}^-} (\mathbf{b}\cdot\mathbf{n})(u-u_h)\,\mathcal{E}(v_h)\,\mathrm{d}s - \sum_{T\in\mathcal{T}}\int_{\partial_- T\setminus\partial\Omega} (\mathbf{b}\cdot\mathbf{n})\lfloor u-u_h\rfloor\mathcal{E}(v_h)\,\mathrm{d}s \\
&\quad + s_h^{a,c}(u-u_h, v_h) + s_h^b(u-u_h, v_h) = 0.
\end{aligned} \tag{6.14}$$

Proof. To begin, from (2.1), (2.3), the consistency of the method yields

$$\begin{aligned}
&\int_\Omega a\nabla u\cdot\nabla\mathcal{E}(v_h) + \mathbf{b}\cdot\nabla u\,\mathcal{E}(v_h) + cu\,\mathcal{E}(v_h)\,\mathrm{d}x - \int_{\Gamma_{\mathrm{D}}} a\nabla u\cdot\mathbf{n}\,\mathcal{E}(v_h)\,\mathrm{d}s \\
&\qquad = \int_\Omega f\mathcal{E}(v_h)\,\mathrm{d}x + \int_{\Gamma_{\mathrm{N}}} g_{\mathrm{N}}\mathcal{E}(v_h)\,\mathrm{d}s \quad \forall v_h \in V_h^r.
\end{aligned} \tag{6.15}$$

Noting that $\mathbf{n}^T a\mathbf{n} = 0$ implies $\mathbf{n}^T a = \mathbf{0}^T$ for a satisfying (2.2). Moreover, the regularity of u [35] allows to also conclude

$$s_h^{a,c}(u, v_h) = s_h^b(u, v_h) = 0 \quad \text{for} \quad v_h \in V_h^r,$$

Subtracting now (5.3) from (6.15) already yields the result. □

Lemma 6.4. *Suppose the assumptions of Theorem 6.1 hold and let u and u_h satisfy the assumptions of Proposition 6.3. In addition, let $\Pi : L^2(\Omega) \to V_h^r$ denote the orthogonal L^2-projection operator onto the polytopic finite element space V_h^r. Upon considering the splitting $u - u_h = (u - \Pi u) - (u_h - \Pi u) =: \eta - \xi$, we have*

$$|B(\xi, v_h)| \leq F(\eta)|||v_h|||_s, \tag{6.16}$$

where

$$
\begin{aligned}
F(\eta)^2 = {} & 2\left\|\sqrt{a}\nabla\left(u-\mathcal{E}\left(\Pi u\right)\right)\right\|^2 + 8C(r)/\sigma_{b,1}\left\|\sqrt{\mathbf{h}}\mathbf{b}\cdot\nabla_h\eta\right\|^2 + 8C(r)\beta/\sigma_{b,1}\left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}\eta\right\|^2_{\Gamma_{\mathrm{D}}^-} \\
& + 8C(r)\beta/\sigma_{b,1}\left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}\lfloor\eta\rfloor\right\|^2_{\Gamma_{\mathrm{int}}} + 2\left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}\eta\right\|^2_{\Gamma_{\mathrm{N}}^+} + 2\left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}\eta^-\right\|^2_{\Gamma_{\mathrm{int}}} \\
& + \left(2\left\|\mathbf{b}\right\|^2_{W^{1,\infty}(\Omega)}/\gamma_0^2\right)\left\|\sqrt{c_0}\eta\right\|^2 + \left(1+2\gamma^2/\gamma_0+8\gamma^2C(r)\mathbf{h}/\sigma_{b,1}\right)\left\|u-\mathcal{E}\left(\Pi u\right)\right\|^2 \\
& + 2\left\|a\nabla\left(u-\mathcal{E}\left(\Pi u\right)\right)\cdot\mathbf{n}\sigma_{\mathrm{D}}^{-1/2}\right\|^2_{\Gamma_{\mathrm{D}}} + 2C(r)\left\|\sqrt{\sigma_{\mathrm{D}}}\left(u-\mathcal{E}\left(\Pi u\right)\right)\right\|^2_{\Gamma_{\mathrm{D}}} \\
& + 2s_h^{a,c}(\eta,\eta) + 2s_h^b(\eta,\eta).
\end{aligned}
$$

Proof. Galerkin orthogonality (6.14), for any $v_h \in V_h^r$, gives

$$
\begin{aligned}
& \int_\Omega \left(a\nabla\left(u-\mathcal{E}\left(\Pi u\right)\right)\cdot\nabla\mathcal{E}\left(v_h\right) + \mathbf{b}\cdot\nabla_h\eta\mathcal{E}\left(v_h\right) + c\left(u-\mathcal{E}\left(\Pi u\right)\right)\mathcal{E}\left(v_h\right)\right)\mathrm{d}x \\
& \quad - \int_{\Gamma_{\mathrm{D}}}\left(a\nabla\left(u-\mathcal{E}\left(\Pi u\right)\right)\cdot\mathbf{n}\,\mathcal{E}\left(v_h\right) + a\nabla\mathcal{E}\left(v_h\right)\cdot\mathbf{n}\,\left(u-\mathcal{E}\left(\Pi u\right)\right) - \sigma_{\mathrm{D}}\left(u-\mathcal{E}\left(\Pi u\right)\right)\mathcal{E}\left(v_h\right)\right)\mathrm{d}s \\
& \quad - \int_{\Gamma_{\mathrm{D}}^-}\left(\mathbf{b}\cdot\mathbf{n}\right)\eta\mathcal{E}\left(v_h\right)\mathrm{d}s - \sum_{T\in\mathcal{T}}\int_{\partial_- T\setminus\partial\Omega}\left(\mathbf{b}\cdot\mathbf{n}\right)\lfloor\eta\rfloor\mathcal{E}\left(v_h\right)\mathrm{d}s + s_h^{a,c}(\eta,v_h) + s_h^b\left(\eta,v_h\right) \\
& = \int_\Omega\left(a\nabla\mathcal{E}(\xi)\cdot\nabla\mathcal{E}\left(v_h\right) + \mathbf{b}\cdot\nabla_h\xi\mathcal{E}\left(v_h\right) + c\mathcal{E}(\xi)\mathcal{E}\left(v_h\right)\right)\mathrm{d}x \\
& \quad - \int_{\Gamma_{\mathrm{D}}}\left(a\nabla\mathcal{E}(\xi)\cdot\mathbf{n}\,\mathcal{E}\left(v_h\right) + a\nabla\mathcal{E}\left(v_h\right)\cdot\mathbf{n}\,\mathcal{E}(\xi) - \sigma_{\mathrm{D}}\mathcal{E}(\xi)\mathcal{E}\left(v_h\right)\right)\mathrm{d}s \\
& \quad - \int_{\Gamma_{\mathrm{D}}^-}\left(\mathbf{b}\cdot\mathbf{n}\right)\xi\mathcal{E}\left(v_h\right)\mathrm{d}s - \sum_{T\in\mathcal{T}}\int_{\partial_- T\setminus\partial\Omega}\left(\mathbf{b}\cdot\mathbf{n}\right)\lfloor\xi\rfloor\mathcal{E}\left(v_h\right)\mathrm{d}s + s_h^{a,c}\left(\xi,v_h\right) + s_h^b\left(\xi,v_h\right).
\end{aligned}
\tag{6.17}
$$

Now examining the terms appearing on the left hand side of (6.17) involving $\mathbf{b}$ we see that

$$
\begin{aligned}
& \int_\Omega \mathbf{b}\cdot\nabla_h\eta\mathcal{E}\left(v_h\right)\,\mathrm{d}x - \int_{\Gamma_{\mathrm{D}}^-}\left(\mathbf{b}\cdot\mathbf{n}\right)\eta\mathcal{E}\left(v_h\right)\mathrm{d}s - \sum_{T\in\mathcal{T}}\int_{\partial_- T\setminus\partial\Omega}\left(\mathbf{b}\cdot\mathbf{n}\right)\lfloor\eta\rfloor\mathcal{E}\left(v_h\right)\mathrm{d}s \\
& \quad = \int_\Omega \mathbf{b}\cdot\nabla_h\eta\left(\mathcal{E}\left(v_h\right)-v_h\right)\,\mathrm{d}x - \int_{\Gamma_{\mathrm{D}}^-}\left(\mathbf{b}\cdot\mathbf{n}\right)\eta\left(\mathcal{E}\left(v_h\right)-v_h\right)\mathrm{d}s \\
& \quad\quad - \sum_{T\in\mathcal{T}}\int_{\partial_- T\setminus\partial\Omega}\left(\mathbf{b}\cdot\mathbf{n}\right)\lfloor\eta\rfloor\left(\mathcal{E}\left(v_h\right)-v_h\right)^+\mathrm{d}s - \int_\Omega\left(\mathbf{b}\cdot\nabla_h v_h + (\nabla\cdot\mathbf{b})v_h\right)\eta\,\mathrm{d}x \\
& \quad\quad + \int_{\Gamma_{\mathrm{N}}^+}\left(\mathbf{b}\cdot\mathbf{n}\right)\eta v_h\mathrm{d}s + \sum_{T\in\mathcal{T}}\int_{\partial_- T\setminus\partial\Omega}\left(\mathbf{b}\cdot\mathbf{n}\right)\lfloor v_h\rfloor\eta^-\,\mathrm{d}s.
\end{aligned}
\tag{6.18}
$$

Recalling that $\mathbf{b}$ satisfies assumption (5.2), by using the orthogonality of Π, we immediately have $\int_\Omega \mathbf{b}\cdot\nabla_h v_h\eta\,\mathrm{d}x = 0$. Using Lemma 3.1 and the Cauchy–Schwarz inequality, we have from (6.18)

$$\begin{aligned}
&\int_\Omega \mathbf{b}\cdot\nabla_h\eta\mathcal{E}\left(v_h\right)\,\mathrm{d}x-\int_{\Gamma_{\mathrm{D}}^-}\left(\mathbf{b}\cdot\mathbf{n}\right)\eta\mathcal{E}\left(v_h\right)\mathrm{d}s-\sum_{T\in\mathcal{T}}\int_{\partial_- T\setminus\partial\Omega}\left(\mathbf{b}\cdot\mathbf{n}\right)\lfloor\eta\rfloor\mathcal{E}\left(v_h\right)\mathrm{d}s\\
&\quad\leq C(r)\left\|\sqrt{\mathbf{h}}\mathbf{b}\cdot\nabla_h\eta\right\|\left\|[v_h]\right\|_{\Gamma_{\mathrm{int}}}+C(r)\sqrt{\beta}\left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}\eta\right\|_{\Gamma_{\mathrm{D}}^-}\left\|[v_h]\right\|_{\Gamma_{\mathrm{int}}}\\
&\quad+C(r)\sqrt{\beta}\left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}\lfloor\eta\rfloor\right\|_{\Gamma_{\mathrm{int}}}\left\|[v_h]\right\|_{\Gamma_{\mathrm{int}}}+\|\mathbf{b}\|_{W^{1,\infty}(\Omega)}\gamma_0^{-1}\left\|\sqrt{c_0}v_h\right\|\left\|\sqrt{c_0}\eta\right\|\\
&\quad+\left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}\eta\right\|_{\Gamma_{\mathrm{N}}^+}\left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}v_h\right\|_{\Gamma_{\mathrm{N}}^+}+\left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}\eta^-\right\|_{\Gamma_{\mathrm{int}}}\left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}\lfloor v_h\rfloor\right\|_{\Gamma_{\mathrm{int}}}.
\end{aligned}\tag{6.19}$$

Now making use of analogous arguments for the reaction term, the triangle inequality and Lemma 3.1, we have

$$\begin{aligned}
\int_\Omega c\left(u-\mathcal{E}\left(\Pi u\right)\right)\mathcal{E}\left(v_h\right)\,\mathrm{d}x&\leq\gamma\left\|u-\mathcal{E}\left(\Pi u\right)\right\|\left(\|v_h\|+\left\|\mathcal{E}\left(v_h\right)-v_h\right\|\right)\\
&\leq\gamma\gamma_0^{-1/2}\left\|u-\mathcal{E}\left(\Pi u\right)\right\|\left\|\sqrt{c_0}v_h\right\|+\gamma C(r)\mathbf{h}^{1/2}\left\|u-\mathcal{E}\left(\Pi u\right)\right\|\left\|[v_h]\right\|_{\Gamma_{\mathrm{int}}}.
\end{aligned}\tag{6.20}$$

As for the Dirichlet boundary, assumption (5.1) together with an inverse inequality, result to

$$\begin{aligned}
\int_{\Gamma_{\mathrm{D}}}a\nabla\mathcal{E}\left(v_h\right)\cdot\mathbf{n}\left(u-\mathcal{E}\left(\Pi u\right)\right)\mathrm{d}s&\leq\left\|\sigma_{\mathrm{D}}^{-1/2}a\nabla\mathcal{E}\left(v_h\right)\cdot\mathbf{n}\right\|_{\Gamma_{\mathrm{D}}}\left\|\sigma_{\mathrm{D}}^{1/2}\left(u-\mathcal{E}\left(\Pi u\right)\right)\right\|_{\Gamma_{\mathrm{D}}}\\
&\leq C(r)\left\|\sigma_{\mathrm{D}}^{1/2}\left(u-\mathcal{E}\left(\Pi u\right)\right)\right\|_{\Gamma_{\mathrm{D}}}\left\|\sqrt{a}\nabla\mathcal{E}\left(v_h\right)\right\|.
\end{aligned}\tag{6.21}$$

The result then follows by combining (6.17)–(6.21), thereby concluding the proof. □

Theorem 6.5. *Suppose that the assumptions of Theorem 6.1 hold and let u and u_h satisfy the assumptions of Proposition 6.3. Suppose further that $u|_T\in H^k(T), T\in\mathcal{T}$, for some $k\geq 2$. Recalling $e=u-u_h$, we have the following a priori error bound:*

$$\begin{aligned}
&\left\|\sqrt{a}\nabla e\right\|^2+\left\|\sqrt{\sigma_{\mathrm{D}}}e\right\|_{\Gamma_{\mathrm{D}}}^2+\left\|u-u_h\right\|_{\mathrm{b}}^2+\left\|\sqrt{\delta\mathbf{h}}\left(\mathbf{b}\cdot\nabla_h\left(u-u_h\right)\right)\right\|^2\\
&\qquad+s_h^{a,c}\left(u-u_h,u-u_h\right)+s_h^b\left(u-u_h,u-u_h\right)\leq C\sum_{T\in\mathcal{T}}\left(\mathcal{D}_T+\mathcal{C}_T\right)\mathbf{h}^{2l-2}|u|_{l,T}^2,
\end{aligned}\tag{6.22}$$

where

$$\mathcal{D}_T=\alpha+\alpha^2\mathbf{h}^{-1}\sigma_{\mathrm{D}}^{-1}+h\sigma_{\mathrm{D}},\tag{6.23}$$

and

$$\begin{aligned}
\mathcal{C}_T=&\beta^2\mathbf{h}/\sigma_{b,1}+\beta\mathbf{h}+\left(\gamma^2/\gamma_0\right)\mathbf{h}^2+\left(\|\mathbf{b}\|_{W^{1,\infty}(\Omega)}^2\left\|c_0\right\|_{L^\infty(\Omega)}/\gamma_0^2\right)\mathbf{h}^2+\left\|c_0\right\|_{L^\infty(\Omega)}\mathbf{h}^2\\
&+\gamma^2C(r)\mathbf{h}^3/\sigma_{b,1}+\delta\beta^2\mathbf{h}+\sigma_{a,c}\mathbf{h}+\sigma_{b,1}\mathbf{h}+\sigma_{b,2}\beta^2\mathbf{h},
\end{aligned}\tag{6.24}$$

for $l=\min\{k,r+1\}$, with C positive constant, independent of u and of $\mathbf{h}$.

Proof. Using the notation of the proof of Lemma 6.4, triangle inequality implies

$$\begin{aligned}
&\left\|\sqrt{a}\nabla e\right\|^2 + \left\|\sqrt{c_0}\,(u-u_h)\right\|^2 + \left\|\sqrt{\sigma_{\mathrm{D}}}e\right\|_{\Gamma_{\mathrm{D}}}^2 + \frac{1}{2}\left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}\,(u-u_h)\right\|_{\Gamma_{\mathrm{D}}^-}^2 \\
&\quad + \frac{1}{2}\left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}\lfloor u-u_h\rfloor\right\|_{\Gamma_{\mathrm{int}}}^2 + \frac{1}{2}\left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}\,(u-u_h)\right\|_{\Gamma_{\mathrm{N}}^+}^2 + \left\|\sqrt{\delta\mathbf{h}}\,(\mathbf{b}\cdot\nabla_h\,(u-u_h))\right\|^2 \\
&\quad + s_h^{a,c}\,(u-u_h,u-u_h) + s_h^b\,(u-u_h,u-u_h) \\
&\leq \Big(2\left\|\sqrt{a}\nabla\,(u-\mathcal{E}\,(\Pi u))\right\|^2 + 2\left\|\sqrt{c_0}\eta\right\|^2 + 2\left\|\sqrt{\sigma_{\mathrm{D}}}\,(u-\mathcal{E}\,(\Pi u))\right\|_{\Gamma_{\mathrm{D}}}^2 \\
&\quad + \left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}\eta\right\|_{\Gamma_{\mathrm{D}}^-}^2 + \left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}\lfloor\eta\rfloor\right\|_{\Gamma_{\mathrm{int}}}^2 + \left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}\eta\right\|_{\Gamma_{\mathrm{N}}^+}^2 + 2\left\|\sqrt{\delta\mathbf{h}}\,(\mathbf{b}\cdot\nabla_h\eta)\right\|^2 \\
&\quad + 2s_h^{a,c}(\eta,\eta) + 2s_h^b(\eta,\eta)\Big) + \left(2|||\xi|||_s^2\right) =: \mathrm{I}+\mathrm{II}.
\end{aligned} \tag{6.25}$$

The optimal approximation properties of L^2-orthogonal projection operator, combined with lemma (3.1), yield

$$\mathrm{I} \leq C\sum_{T\in\mathcal{T}}\left(\alpha+\sigma_{\mathrm{D}}\mathbf{h}+\mathbf{h}\left(\mathbf{h}\left\|c_0\right\|_{L^\infty(\Omega)}+\beta+\delta\beta^2+\sigma_{a,c}+\sigma_{b,1}+\sigma_{b,2}\beta^2\right)\right)\mathbf{h}^{2l-2}|u|_{l,T}^2. \tag{6.26}$$

Next, the inf-sup condition given in Theorem 6.1 along with Lemma 6.4 give

$$\begin{aligned}
\mathrm{II} \leq{}& \frac{4}{\Lambda^2}\Big(2\left\|\sqrt{a}\nabla\,(u-\mathcal{E}\,(\Pi u))\right\|^2 + C(r)/\sigma_{b,1}\left\|\sqrt{\mathbf{h}}\mathbf{b}\cdot\nabla_h\eta\right\|^2 + C(r)\beta/\sigma_{b,1}\left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}\eta\right\|_{\Gamma_{\mathrm{D}}^-}^2 \\
&+ C(r)\beta/\sigma_{b,1}\left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}\lfloor\eta\rfloor\right\|_{\Gamma_{\mathrm{int}}}^2 + \left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}\eta\right\|_{\Gamma_{\mathrm{N}}^+}^2 + \left\|\sqrt{|\mathbf{b}\cdot\mathbf{n}|}\eta^-\right\|_{\Gamma_{\mathrm{int}}}^2 \\
&+ \gamma^2(\gamma_0^{-1}+C(r)\mathbf{h}/\sigma_{b,1})\left\|u-\mathcal{E}\,(\Pi u)\right\|^2 + \left\|\mathbf{b}\right\|_{W^{1,\infty}(\Omega)}^2/\gamma_0^2\left\|\sqrt{c_0}\eta\right\|^2 \\
&+ \left\|a\nabla\,(u-\mathcal{E}\,(\Pi u))\,\sigma_{\mathrm{D}}^{-1/2}\right\|_{\Gamma_{\mathrm{D}}}^2 + C(r)\left\|\sqrt{\sigma_{\mathrm{D}}}\,(u-\mathcal{E}\,(\Pi u))\right\|_{\Gamma_{\mathrm{D}}}^2 + s_h^{a,c}(\eta,\eta) + s_h^b(\eta,\eta)\Big) \\
\leq{}& C\sum_{T\in\mathcal{T}}\Big(\alpha+\beta^2\mathbf{h}/\sigma_{b,1}+\beta\mathbf{h}+\left(\gamma^2/\gamma_0\right)\mathbf{h}^2 + \left\|\mathbf{b}\right\|_{W^{1,\infty}(\Omega)}^2\left\|c_0\right\|_{L^\infty(\Omega)}/\gamma_0^2\mathbf{h}^2 \\
&+ \gamma^2 C(r)\mathbf{h}^3/\sigma_{b,1}+\alpha^2\mathbf{h}^{-1}\sigma_{\mathrm{D}}^{-1}+\mathbf{h}\sigma_{\mathrm{D}}+\sigma_{a,c}\mathbf{h}+\sigma_{b,1}\mathbf{h}+\sigma_{b,2}\beta^2\mathbf{h}\Big)\mathbf{h}^{2l-2}|u|_{l,T}^2.
\end{aligned} \tag{6.27}$$

The result then follows by combining the bounds (6.26) and (6.27). □

Remark 6.6. We note that the above *a priori* error bound for R-FEM (6.22) is optimal with respect to the mesh-size h. If $\mathbf{b}=\mathbf{0}$, the error bound (6.22) reduces to

$$\left\|\sqrt{a}\nabla\,(u-\mathcal{E}\,(u_h))\right\|^2 + \left\|\sqrt{\sigma_{\mathrm{D}}}\,(u-\mathcal{E}\,(u_h))\right\|_{\Gamma_{\mathrm{D}}}^2 + s_h^{a,0}\,(u-u_h,u-u_h) \leq C\mathbf{h}^{2l-2}|u|_l^2.$$

The above error bound is h-optimal and coincides the error bound from [24] which was shown for standard meshes consisting of simplices. On the other hand, if in the pure hyperbolic case with diffusion tensor $a=0$, we deduce the error bound

$$\left\|u-u_h\right\|_{\mathrm{b}}^2 + \left\|\sqrt{\delta\mathbf{h}}\,(\mathbf{b}\cdot\nabla_h\,(u-u_h))\right\|^2 + s_h^{0,c}\,(u-u_h,u-u_h) + s_h^b\,(u-u_h,u-u_h) \leq C\mathbf{h}^{2l-1}|u|_l^2,$$

which is also h-optimal.

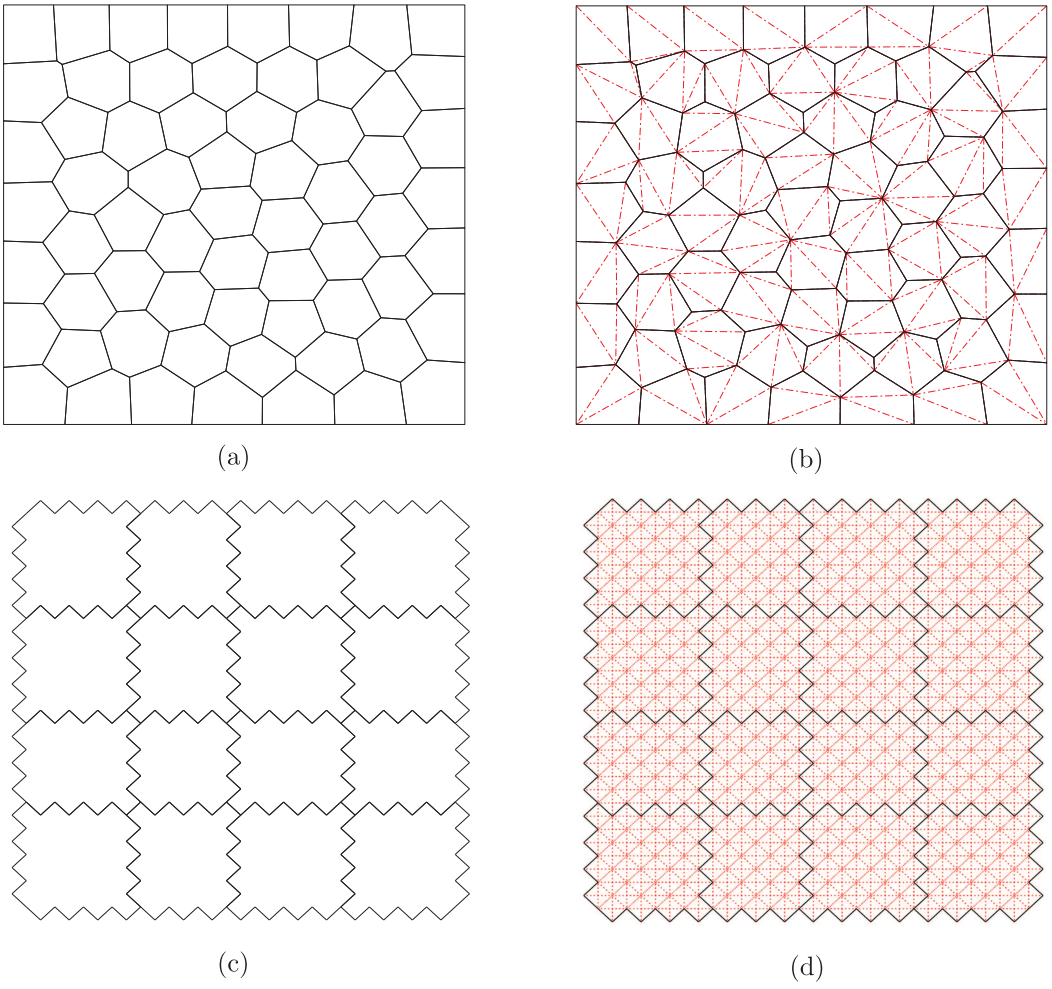
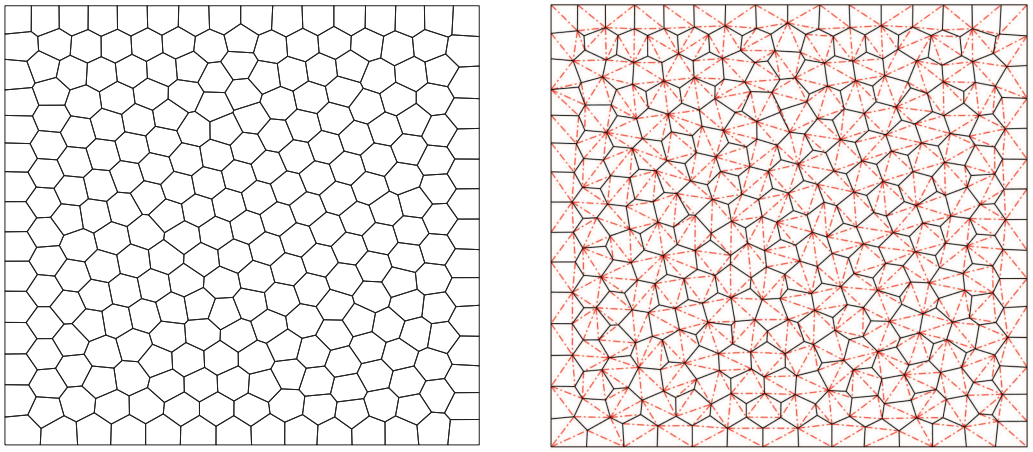


FIGURE 1. Two examples of a polygonal meshes and respective simplicial subdivisions. (a) A polygonal mesh $\mathcal{T}_1$. (b) A simplicial subdivision $\widetilde{\mathcal{T}}_1$. (c) An example of an agglomerated mesh with many tiny faces $\mathcal{T}_2$. (d) A simplicial background mesh $\widetilde{\mathcal{T}}_2$.

7. Numerical experiments

We shall now investigate numerically the asymptotic behaviour of the proposed R-FEM method on general polygonal meshes. We first introduce a sequence of polygonal meshes, indexed by their element size, together with the simplicial sub-meshes used; see Figure 1a for an example or Figure 2 for a refinement of the latter. We point out that the sub-triangulations used do not introduce any new points in the interior of the polygonal mesh to keep the number of degrees of freedom in the triangulated sub-meshes to a minimum. The numerical quadratures are done over the simplicial sub-meshes, where the recovery operator from Lemma 3.1 is applied. As expected from the theory, we have numerically observed that, increasing the number of degrees of freedom in the sub-meshes as a proportion of the polytopic meshes, does not increase the order of the method and only improves the accuracy marginally. All the polygonal meshes are generated by PolyMesher [42]. Unless clearly stated, the R-FEM solution u_h is computed by (5.4) with the following choices of stabilisation parameters C_σ appearing in σ_{D}, $\sigma_{a,c}$ in (5.5), $\sigma_{b,1}$ and $\sigma_{b,2}$ in (5.6), all with value equal to 10.



(a) (b)

FIGURE 2. An example of a polytopic mesh $\mathcal{T}$ which is a refinement of the mesh from Figure 1. (a) $\mathcal{T}$ with 256 polygons. (b) 966-triangle sub-mesh.

7.1. Example 1: A first order hyperbolic problem

Let $\Omega := (0,1)^2$, and choose

$$a \equiv 0, \quad \mathbf{b} = \left(2 - y^2, 2 - x\right), \quad c = 1 + (1+x)(1+y)^2; \tag{7.1}$$

the forcing function f is selected so that the analytical solution to (2.1) is given by

$$u(x,y) = 1 + \sin\left(\pi(1+x)(1+y)^2/8\right), \tag{7.2}$$

cf., [11,31].

We examine the convergence behaviour of the R-FEM with respect to h-refinement, with fixed polynomial r, for $r = 1,\ldots,4$. In Figure 3 the error in two different norms, against the square root of the number of degrees of freedom in the underlying finite element space V_h^r is given. The slope of the convergence rate shown is the slope of the last line segment for each convergence line. We observe that $\|u - u_h\|_\Omega$ and $\|u - u_h\|_{\mathrm{b}}$ converge to zero at the optimal rates $\mathcal{O}(h^{r+1})$ and at least $\mathcal{O}\left(h^{r+\frac{1}{2}}\right)$, respectively, as the mesh size h tends to zero for each fixed r. The latter results agree with the result (6.22) in Theorem 6.5. Notice that the $\|u - u_h\|_{\mathrm{b}}$ appears to superconverge for r even. It has been observed in the literature of hyperbolic conservation laws that often, numerically, one may observe super optimal convergence rates, higher than the provable $O\left(h^{r+1/2}\right)$, for even polynomial degree discontinuous Galerkin methods [17,27].

7.2. Example 2: A nonsymmetric elliptic problem

Let $\Omega := (0,1)^2$, and choose

$$a \equiv 1, \quad \mathbf{b} = (1-y, 1-x), \quad c \equiv 2; \tag{7.3}$$

the forcing function f is selected so that the analytical solution to (2.1) is given by

$$u(x,y) = \sin(\pi x)\sin(\pi y). \tag{7.4}$$

We examine the convergence behaviour of the R-FEM with respect to h-refinement on quasiuniform polygonal meshes, with fixed polynomial r, for $r = 1,\ldots,4$. In Figure 4 we plot the error, in terms of the L^2-norm

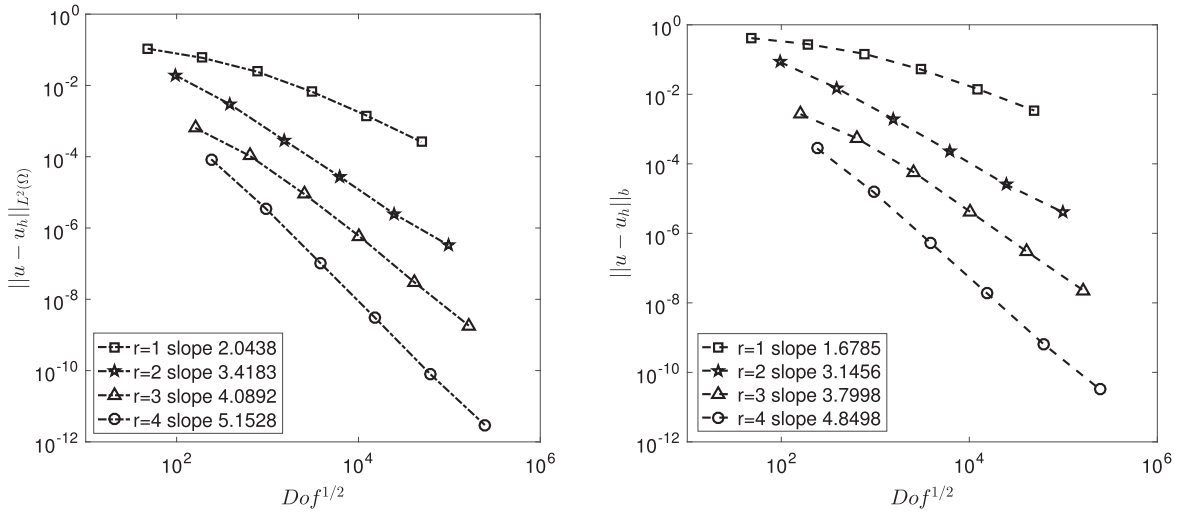


FIGURE 3. Example 1. Error against numerical degrees of freedom (Dof). Here we examine the convergence of the R-FEM under h-refinement for polynomial degrees $r = 1, 2, 3, 4$. Notice that $\|u - u_h\|_{L^2(\Omega)} = O\left(h^{r+1}\right)$, which is optimal. In addition, $\|u - u_h\|_{\mathrm{b}}$ appears to converge faster than the analysis of Theorem 6.5 suggests for even polynomial degrees.

and the (broken) H^1-seminorm for both discontinuous R-FEM approximation u_h and the conforming R-FEM approximation $\mathcal{E}(u_h)$, against the square root of the number of degrees of freedom in the underlying finite element space V_h^r. We observe that $\|u - u_h\|$ and $|u - u_h|_{H^1(\Omega,\mathcal{T})}$ converge to zero at the optimal rates $\mathcal{O}\left(h^{r+1}\right)$ and $\mathcal{O}\left(h^r\right)$, respectively, as the mesh size h tends to zero for each fixed r. Moreover, the difference between the R-FEM solutions u_h and $\mathcal{E}(u_h)$ is marginal. The results are in accordance with Theorem 6.5.

7.3. Example 3: A advection-dominated elliptic problem

We now investigate the numerical stability of the R-FEM through a series of advection-dominated elliptic problems. Let $\Omega := (0,1)^2$, and choose

$$a \equiv \epsilon, \quad \mathbf{b} = (1,1), \quad c \equiv 0; \tag{7.5}$$

together with a forcing function $f = 1$ and homogeneous Dirichlet boundary conditions. This example is known to admit boundary layers in the vicinity of the right and top boundaries $x = 1$ and $y = 1$ when $\epsilon \ll 1$. We investigate the stability of R-FEM as $\epsilon \to 0$ on a fixed, relatively coarse mesh which is insufficient to resolve the layer. More specifically, we consider a fixed mesh consisting of 1024 polygons over the domain, we choose $r = 1$ and take $\epsilon = 10^{-2}, 10^{-4}, 10^{-6}$. In Figure 5 we plot the numerical solutions u_h and $\mathcal{E}(u_h)$. We observe that for $\epsilon = 10^{-2}$, the mesh is fine enough to resolve the layer and, hence, both u_h and $\mathcal{E}(u_h)$ are stable. For $\epsilon = 10^{-4}$, the mesh is no longer fine enough to resolve the layer. However, the solutions are still stable in the sense that neither solution admits non-physical oscillations near the boundary. In the case $\epsilon = 10^{-6}$, the mesh is too coarse to resolve the layer. Both u_h and $\mathcal{E}(u_h)$ appear to be stable in this case. Moreover, the solutions are very close to the solution for the pure hyperbolic problem with inflow boundary satisfying Dirichlet boundary conditions. This is expected as the boundary conditions have been imposed in a weak fashion for the numerical method to be valid in the hyperbolic limit $\epsilon = 0$ also, in the spirit of the classical discontinuous Galerkin methods.

Finally, we make a comparison between the proposed R-FEM and a post-processed solution $\mathcal{E}(u_{DG})$ of an interior penalty dG solution u_{DG} from [10] over the same 1024-polygon mesh. In Figure 6, we provide the numerical solution with the same $\epsilon = 10^{-4}$ and linear elements. As it can been seen, the two solutions have

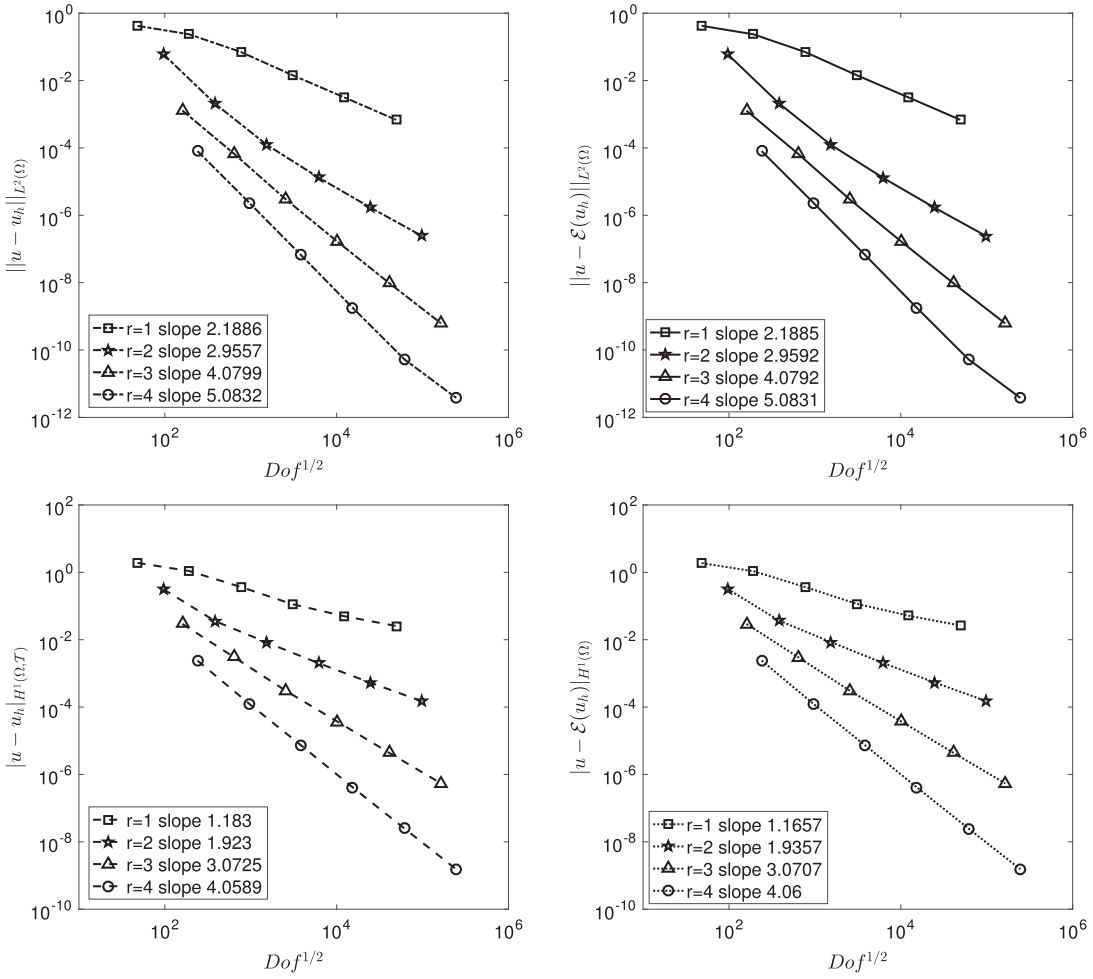


FIGURE 4. Example 2. Convergence of the R-FEM under h-refinement for $r = 1, 2, 3, 4$.

different qualitative properties, with R-FEM appearing to be giving qualitatively similar stabilisation with conforming-type stabilization methods (which may *not* be directly generalizable to polytopic meshes), rather than the typical DG profile of exponentially diminished oscillations away from the layer and strongly oscillatory layer behaviour.

7.4. Example 4: A mixed-type problem

We now consider a partial differential equation with nonnegative characteristic form of mixed type. To this end, we let $\Omega = (-1, 1)^2$, and consider the PDE problem:

$$\begin{cases} -x^2 u_{yy} + u_x + u = 0, & \text{for} \quad -1 \le x \le 1, y > 0, \\ u_x + u = 0, & \text{for} \quad -1 \le x \le 1, y \le 0, \end{cases} \tag{7.6}$$

with analytical solution:

$$u(x,y) = \begin{cases} \sin\left(\frac{1}{2}\pi(1+y)\right) \exp\left(-\left(x + \frac{\pi^2 x^3}{12}\right)\right), & \text{for} \quad -1 \le x \le 1, y > 0, \\ \sin\left(\frac{1}{2}\pi(1+y)\right) \exp(-x), & \text{for} \quad -1 \le x \le 1, y \le 0, \end{cases} \tag{7.7}$$

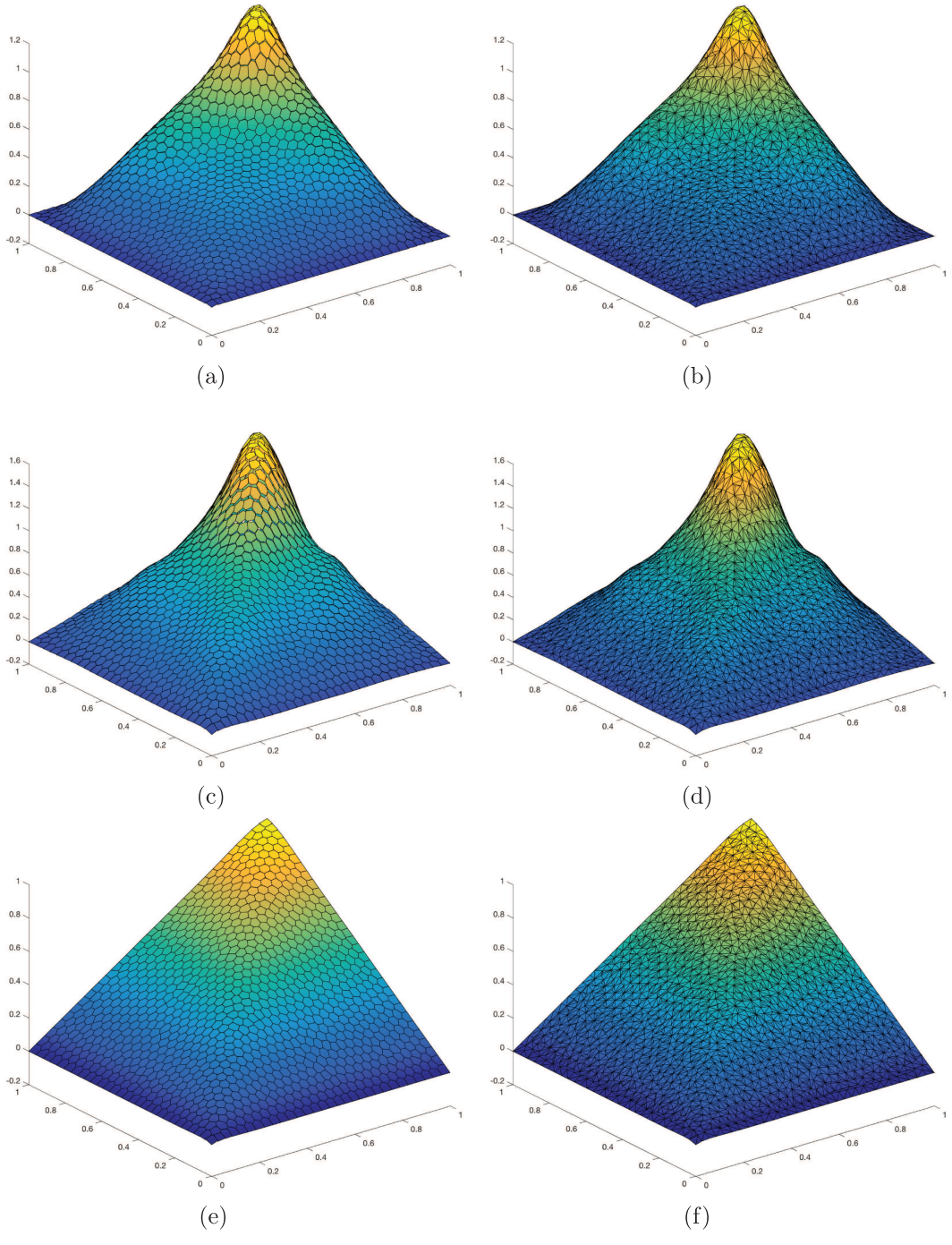


FIGURE 5. Example 3. R-FEM solutions for a mesh consisting of 1024 polygonal elements and $r = 1$. (a) u_h with $\epsilon = 10^{-2}$. (b) $\mathcal{E}(u_h)$ with $\epsilon = 10^{-2}$. (c) u_h with $\epsilon = 10^{-4}$. (d) $\mathcal{E}(u_h)$ with $\epsilon = 10^{-4}$. (e) u_h with $\epsilon = 10^{-6}$. (f) $\mathcal{E}(u_h)$ with $\epsilon = 10^{-6}$.

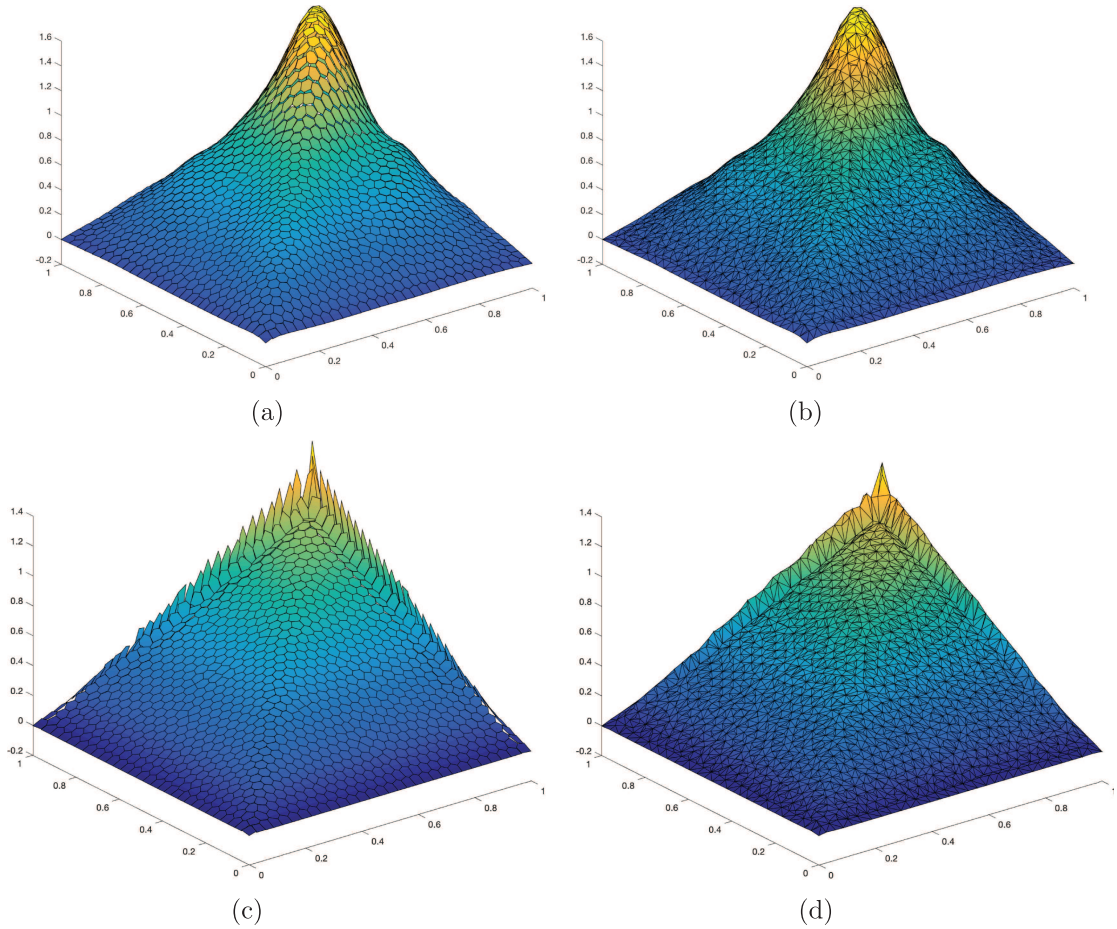
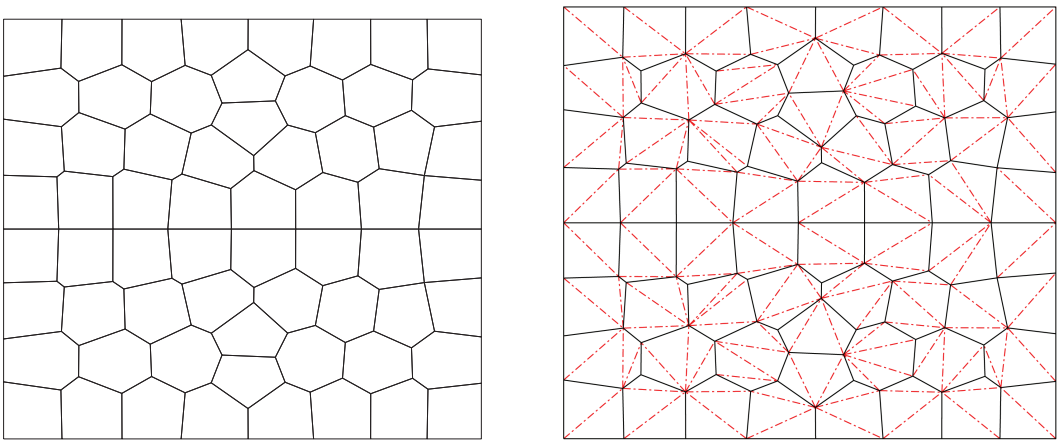


FIGURE 6. Example 3. R-FEM solutions and DG solutions for a mesh consisting of 1024 polygonal elements and $r = 1$. (a) u_h with $\epsilon = 10^{-4}$. (b) $\mathcal{E}(u_h)$ with $\epsilon = 10^{-4}$. (c) u_{DG} with $\epsilon = 10^{-4}$. (d) $\mathcal{E}(u_{\mathrm{DG}})$ with $\epsilon = 10^{-4}$.

cf., [11]. This problem is hyperbolic in the region $y \leq 0$ and parabolic for $y > 0$. Notice that, in order to ensure continuity of the normal flux across $y = 0$ where the partial differential equation changes type, the analytical solution has a discontinuity across the line $y = 0$, *cf.*, [31].

We examine the convergence behaviour of the R-FEM with respect to h-refinement, with fixed polynomial r, for $r = 1, \ldots, 4$. To have the opportunity to possibly observe optimal convergence rates, we align the polygonal mesh with the solution's discontinuity; a typical mesh is shown in Figure 7. Also, for this example the recovery operator is constructed in piecewise fashion over the two subdomains. This ensures the conforming R-FEM solution $\mathcal{E}(u_h)$ is able to have a jump discontinuity over the interface where the problem changes type.

In Figure 8 we plot the L^2-norm error, as well as the error in the norm on the left hand-side of (6.22) for R-FEM approximation u_h, against the square root of the number of degrees of freedom in the underlying finite element space V_h^r. Abusing the notation, we use again $|||\cdot|||_s$ to denote the norm on the left hand-side of (6.22). We observe that $|||u - u_h|||_s$ converges to zero at the optimal rates $\mathcal{O}(h^r)$, as the mesh size h tends to zero for each fixed r. These results agree with the result (6.22) in Theorem 6.5. However, the convergence rate for $\|u - u_h\|$ seems to be slightly suboptimal in h. Additionally, we also plot the error in terms of L^2-norm and



(a) (b)

FIGURE 7. An aligned polytopic mesh, $\mathcal{T}$, with 64 polygons and 208 triangle sub-mesh $\widetilde{\mathcal{T}}$. (a) An aligned mesh with 64 polygons. (b) A 208 triangle sub-mesh.

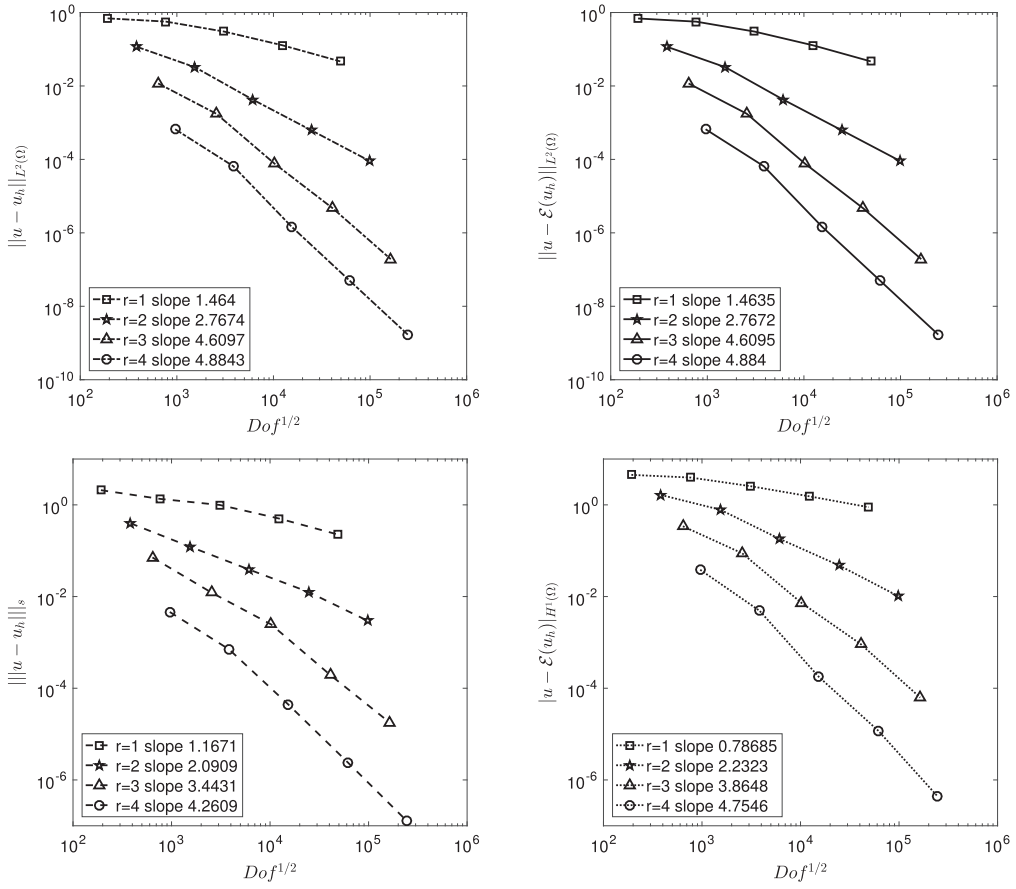
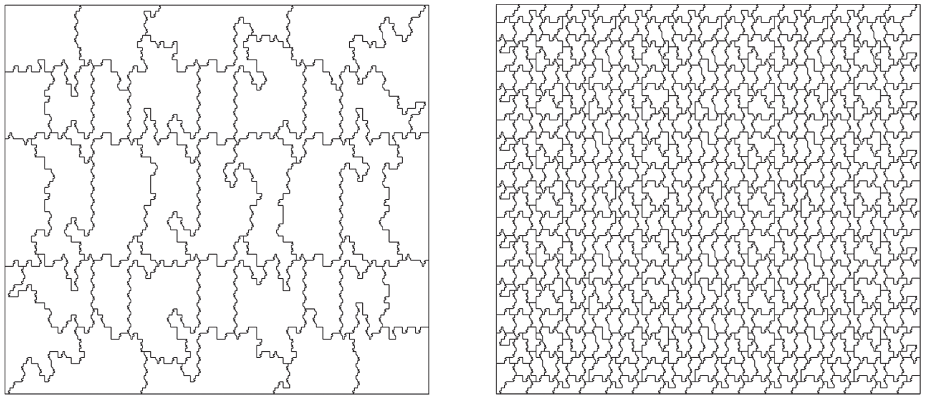


FIGURE 8. Example 4. Convergence of the R-FEM under h-refinement for $r = 1, 2, 3, 4$.



(a) (b)

FIGURE 9. An example of agglomerated meshes $\mathcal{T}$ with a lot of tiny faces. (a) $\mathcal{T}$ with 35 polygons. (b) $\mathcal{T}$ with 548 polygons.

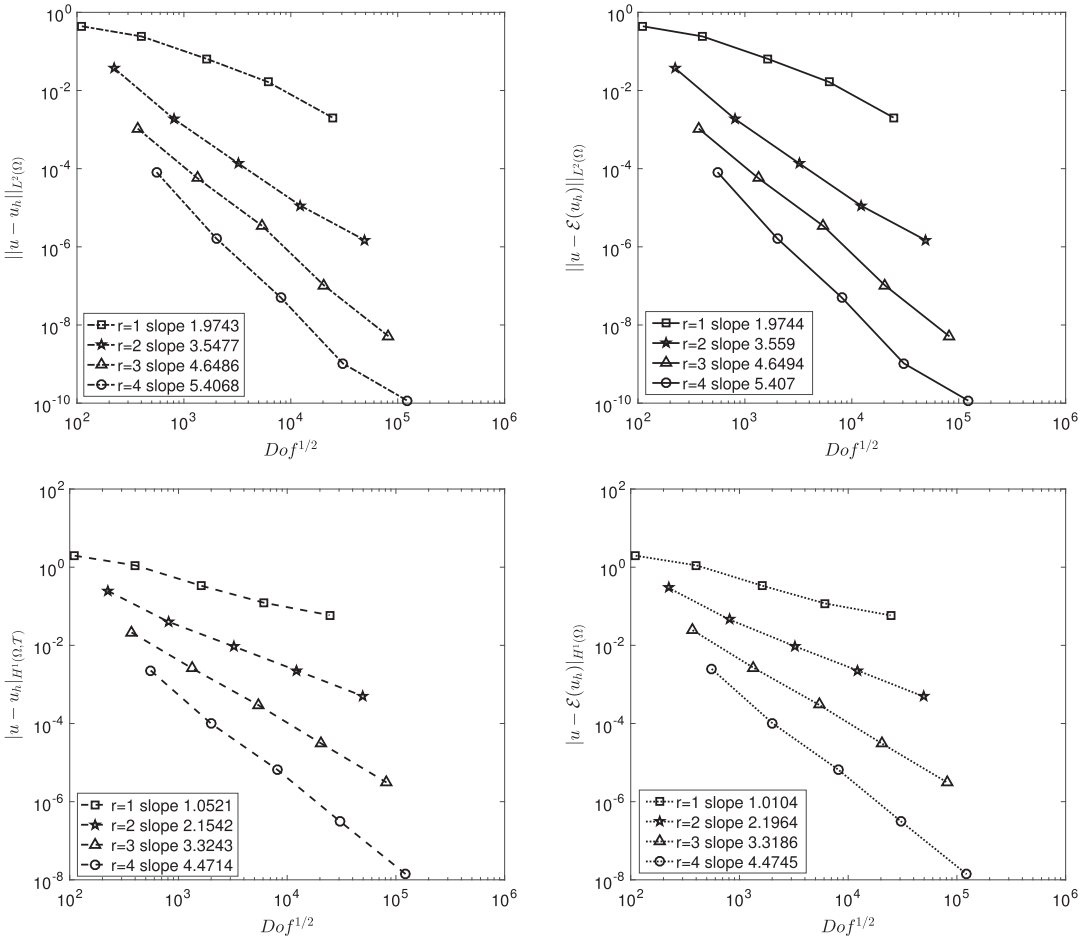


FIGURE 10. Example 5. Convergence of the R-FEM under h-refinement for $r = 1, 2, 3, 4$.

H^1-seminorm for $\mathcal{E}(u_h)$, against the square root of the number of degrees of freedom in the underlying finite element space V_h^r. Here, again, observe that $\|u - \mathcal{E}(u_h)\|$ and $|u - \mathcal{E}(u_h)|_{H^1\Omega}$ converge to zero at a slightly suboptimal rate.

7.5. Example 5: An elliptic problem over agglomerated meshes with many faces per element

Let $\Omega := (0,1)^2$, and choose

$$a \equiv 1, \quad \mathbf{b} = (0,0), \quad c = 0; \tag{7.8}$$

the forcing function f is selected so that the analytical solution to (2.1) is given by

$$u(x,y) = \sin(\pi x)\sin(\pi y). \tag{7.9}$$

In this example, we investigate the convergence behaviour of the R-FEM with respect to h-refinement on agglomerated meshes, with fixed polynomial r, for $r = 1,\ldots,4$. Here the successive fine agglomerated meshes are constructed based on about half a million uniform triangles. In the coarsest level of the agglomerated meshes, each polygon has approximately 300 tiny faces; a typical mesh is shown in Figure 9. In Figure 4 we plot the error, in terms of the L^2-norm and the (broken) H^1-seminorm for both discontinuous R-FEM approximation u_h and the conforming R-FEM approximation $\mathcal{E}(u_h)$, against the square root of the number of degrees of freedom in the underlying finite element space V_h^r. We observe that $\|u - u_h\|$ and $|u - u_h|_{H^1(\Omega,\mathcal{T})}$ converge to zero at the optimal rates $\mathcal{O}\left(h^{r+1}\right)$ and $\mathcal{O}(h^r)$, respectively, as the mesh size h tends to zero for each fixed r. Moreover, the convergence rate seems to be better than the optimal rate possibly because the large number of background meshes take more information to the R-FEM solution. We emphasize that the theoretically analysis in this work does not hold for R-FEM over these agglomerated meshes because of the presence of small edges, but the numerical results are in accordance with Theorem 6.5 (Fig. 10).

8. Conclusion

In this work, we extended the recently developed R-FEM and applied it to PDEs with nonnegative characteristic form. We have defined our scheme over general meshes, consisting of polytopic elements showing that the total degrees of freedom of the R-FEM solution only depends on the number of elements in the polytopic mesh. We have shown the R-FEM is stable for convection-dominated problems, pure hyperbolic problems and problems of mixed classification when the interface is aligned with the polytopic mesh. Numerical experiments have been presented to confirm the theoretical results derived in this paper. As a byproduct of the solution process, we obtain a conforming counterpart of the solution, defined over a subtriangulation of the polytopic mesh. This conforming approximation plays a key role in deriving *a-posteriori* error control for the R-FEM which will be considered in the future work.

Acknowledgements. We wish to express our sincere gratitude Andrea Cangiani (University of Leicester) for his insightful comments on an earlier version of this work. ZD and EHG acknowledge funding by the Leverhulme Trust (grant no. RPG-2015-306) and TP acknowledges funding by the EPSRC (grant no. EP/P000835/1).

References

[1] A.V. Astaneh, F. Fuentes, J. Mora and L. Demkowicz, High-order polygonal discontinuous Petrov–Galerkin (PolyDPG) methods using ultraweak formulations. *Comput. Methods Appl. Mech. Eng.* **332** (2018) 686–711.

[2] G.R. Barrenechea, E.H. Georgoulis and T. Pryer, Recovered mixed finite element methods. In preparation (2020).

[3] L. Beirão da Veiga, F. Brezzi, A. Cangiani, G. Manzini, L. Marini and A. Russo, Basic principles of virtual element methods. *Math. Models Methods Appl. Sci.* **23** (2013) 199–214.

[4] L. Beirão da Veiga, K. Lipnikov and G. Manzini, The Mimetic Finite Difference Method for Elliptic Problems. In: Vol. 11 of *MS&A. Modeling, Simulation and Applications*. Springer, Cham (2014).

[5] L. Beirão da Veiga, C. Lovadina and A. Russo, Stability analysis for the virtual element method. *Math. Models Methods Appl. Sci.* **27** (2017) 2557–2594.

[6] S.C. Brenner and L.-Y. Sung, C^0 interior penalty methods for fourth order elliptic boundary value problems on polygonal domains. *J. Sci. Comput.* **22/23** (2005) 83–118.

[7] S.C. Brenner and L.-Y. Sung, Virtual element methods on meshes with small edges or faces. *Math. Models Methods Appl. Sci.* **28** (2018) 1291–1336.

[8] E. Burman, A unified analysis for conforming and nonconforming stabilized finite element methods using interior penalty. *SIAM J. Numer. Anal.* **43** (2005) 2012–2033.

[9] E. Burman and P. Hansbo, Edge stabilization for Galerkin approximations of convection–diffusion–reaction problems. *Comput. Methods Appl. Mech. Eng.* **193** (2004) 1437–1453.

[10] A. Cangiani, E.H. Georgoulis and P. Houston, *hp*-version discontinuous Galerkin methods on polygonal and polyhedral meshes. *Math. Models Methods Appl. Sci.* **24** (2014) 2009–2041.

[11] A. Cangiani, Z. Dong, E.H. Georgoulis and P. Houston, *hp*-version discontinuous Galerkin methods for advection-diffusion-reaction problems on polytopic meshes. *ESAIM: M2AN* **50** (2016) 699–725.

[12] A. Cangiani, Z. Dong, E.H. Georgoulis and P. Houston, *hp*-Version Discontinuous Galerkin Methods on Polygonal and Polyhedral Meshes. Springer (2017).

[13] A. Cangiani, Z. Dong and E.H. Georgoulis, *hp*-version space-time discontinuous Galerkin methods for parabolic problems on prismatic meshes. *SIAM J. Sci. Comput.* **39** (2017) A1251–A1279.

[14] A. Cangiani, G. Manzini and O.J. Sutton, Conforming and nonconforming virtual element methods for elliptic problems. *IMA J. Numer. Anal.* **37** (2017) 1317–1354.

[15] P.G. Ciarlet, The Finite Element Method for Elliptic Problems. In: Vol. 40 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2002).

[16] P. Clément, Approximation by finite element functions using local regularization. *RAIRO Anal. Numér.* **9** (1975) 77–84.

[17] B. Cockburn, Continuous dependence and error estimation for viscosity methods. *Acta Numer.* **12** (2003) 127–180.

[18] B. Cockburn, J. Gopalakrishnan and R. Lazarov, Unified hybridization of discontinuous Galerkin, mixed, and continuous Galerkin methods for second order elliptic problems. *SIAM J. Numer. Anal.* **47** (2009) 1319–1365.

[19] B. Cockburn, D. Di Pietro and A. Ern, Bridging the hybrid high-order and hybridizable discontinuous Galerkin methods. *ESAIM: M2AN* **50** (2016) 635–650.

[20] D. Di Pietro and A. Ern, Mathematical Aspects of Discontinuous Galerkin Methods. In: Vol. 69 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer, Heidelberg (2012).

[21] D.A. Di Pietro and A. Ern, A hybrid high-order locking-free method for linear elasticity on general meshes. *Comput. Methods Appl. Mech. Eng.* **283** (2015) 1–21.

[22] J. Douglas, Jr. and T. Dupont, Interior penalty procedures for elliptic and parabolic Galerkin methods. In: Vol. 58 of *Lecture Notes in Physics* (1976) 207–216.

[23] E.H. Georgoulis and A. Lasis, A note on the design of *hp*-version interior penalty discontinuous Galerkin finite element methods for degenerate problems. *IMA J. Numer. Anal.* **26** (2006) 381–390.

[24] E.H. Georgoulis and T. Pryer, Recovered finite element methods. *Comput. Methods Appl. Mech. Eng.* **332** (2018) 303–324.

[25] E.H. Georgoulis, P. Houston and J. Virtanen, An a posteriori error indicator for discontinuous Galerkin approximations of fourth-order elliptic problems. *IMA J. Numer. Anal.* **31** (2011) 281–298.

[26] E. Georgoulis, C. Makridakis and T. Pryer, Babuška–Osborn techniques in discontinuous Galerkin methods: L^2-norm error estimates for unstructured meshes. Preprint arXiv:1704.05238 (2017).

[27] J. Giesselmann, C. Makridakis and T. Pryer, A posteriori analysis of discontinuous Galerkin schemes for systems of hyperbolic conservation laws. *SIAM J. Numer. Anal.* **53** (2015) 1280–1303.

[28] W. Hackbusch and S. Sauter, Composite finite elements for the approximation of PDEs on domains with complicated microstructures. *Numer. Math.* **75** (1997) 447–472.

[29] P. Houston and E. Süli, Stabilised *hp*-finite element approximation of partial differential equations with nonnegative characteristic form. *Computing* **66** (2001) 99–119.

[30] P. Houston, C. Schwab and E. Süli, Stabilized *hp*-finite element methods for first-order hyperbolic problems. *SIAM J. Numer. Anal.* **37** (2000) 1618–1643.

[31] P. Houston, C. Schwab and E. Süli, Discontinuous *hp*-finite element methods for advection-diffusion-reaction problems. *SIAM J. Numer. Anal.* **39** (2002) 2133–2163.

[32] O.A. Karakashian and F. Pascal, Convergence of adaptive discontinuous Galerkin approximations of second-order elliptic problems. *SIAM J. Numer. Anal.* **45** (2007) 641–665.

[33] L. Mu, J. Wang and X. Ye, Weak Galerkin finite element methods for the biharmonic equation on polytopal meshes. *Numer. Methods Part. Differ. Equ.* **30** (2014) 1003–1029.

[34] J. Nitsche, über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind. Collection of articles dedicated to Lothar Collatz on his sixtieth birthday. *Abh. Math. Sem. Univ. Hamburg* **36** (1971) 9–15.

[35] O. Oleinik and E. Radkevič, Second Order Equations with Nonnegative Characteristic Form. American Mathematical Society (1973).

[36] P. Oswald, On a BPX-preconditioner for P1 elements. *Computing* **51** (1993) 125–133.

[37] D. Peterseim and S.A. Sauter, The composite mini element-coarse mesh computation of Stokes flows on complicated domains. *SIAM J. Numer. Anal.* **46** (2008) 3181–3206.

[38] W. Reed and T. Hill, Triangular mesh methods for the neutron transport equation. Technical Report LA-UR-73-479, Los Alamos Scientific Laboratory (1973).

[39] S. Rjasanow and S. Weißer, Higher order BEM-based FEM on polygonal meshes. *SIAM J. Numer. Anal.* **50** (2012) 2357–2378.

[40] L.R. Scott and S. Zhang, Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Math. Comput.* **54** (1990) 483–493.

[41] N. Sukumar and A. Tabarraei, Conforming polygonal finite elements. *Int. J. Numer. Methods Eng.* **61** (2004) 2045–2066.

[42] C. Talischi, G. Paulino, A. Pereira and I. Menezes, PolyMesher: a general-purpose mesh generator for polygonal elements written in Matlab. *Struct. Multidisc. Optim.* **45** (2012) 309–328.

[43] A. Veeser and P. Zanotti, Quasi-optimal nonconforming methods for symmetric elliptic problems. III – discontinuous Galerkin and other interior penalty methods. *SIAM J. Numer. Anal.* **56** (2018) 2871–2894.